New perspectives in analyzing aspectual distinctions across languages

Christiane von Stutterheim, Mary Carroll and Wolfgang Klein

1. Introduction

There is a clear distinction between an **event** – the word is taken here in a broad sense[1] – and the **description** of a given event by some speaker on some occasion. A famous murder in the antique world involved two participants, Brutus and Caesar, some activity on the part of the former and some changes that happened to the latter: first, he was alive, and then, he was not alive. These and other features of the event – the time and the place at which it happened and the instrument which Brutus used – are part of reality. A speaker, faced with the task of giving a fairly accurate description of this event, has a number of options. He may or may not choose to take into account some of the entities – persons and objects – that are involved:

(1)   a. Brutus killed Caesar with a dagger.
       b. Brutus killed Caesar.
       c. Caesar died.

He may decide to present this event as on-going, or as completed:

(2)   a. Brutus was killing Caesar with a dagger.
       b. Brutus killed Caesar.

In this case, the specific language, English, provides the speaker with two verbal forms (a "continuous form" and a "simple form") which differentiate between these two options. Other languages use adverbials or periphrastic constructions of a different type to this end. The speaker may present the description in one clause, or distribute it over several clauses or sentences:

---

[1]   It includes all sorts of situations that can occur or obtain in the real or in a fictitious world, including states, processes, actions, or events in a narrower sense, etc.

(3)   a. Brutus killed Caesar.
     b. Brutus decided to kill Caesar for political reasons. He carried out his plan immediately. He took a dagger and stabbed Caesar several times. Caesar died on the stairs of the Curia.

There are many other options in giving a description of one and the same event. The speaker can add all sorts of additional features, from the facial expression of the protagonist to the shape of the instrument involved. But these options are not unrestricted. There are at least four types of more or less strong constraints that influence or determine the speaker's options:

A. A first type of constraint comes with the particular communicative task that the speaker wants to perform. The nature of this task determines the choice of features that are made explicit and the level of detail – the "granularity" – with which the facts are presented (see von Stutterheim 1997).

B. A second type of constraint is given with the lexical and grammatical properties of the particular language used. Some languages, English, for example, force their speakers to indicate the time at which the event occurred, since tense marking on the verb is obligatory. The speaker must therefore locate the event in the past, present or future. Other languages, such as Chinese, leave it to the speaker's discretion to provide information about the "when" of the event. Similarly, some languages have a form that is neutral with respect to "on-going" and "completed"; other languages force their speakers to choose between one of these options, while another set force them to make an aspectual choice for one tense but not for another.

C. A third type of constraint is given with the individual level of proficiency with which the means provided by the language are used: there are highly proficient speakers, and ones who are less so. It is one thing to know what is structurally and lexically possible, and quite another matter to use this potential in solving a particular communicative task. It is not enough to have a tool box, one must also be able to use it. This is particularly clear in the case of language learners who are still on their way to becoming more or fully proficient.

D. Finally, speakers may also be constrained in the choices they make by particular cognitive, cultural or social habits. Although someone could present a certain event in a certain way, he will be unlikely to do so because this may not correspond to the way he or the community to which he belongs would normally relate to it. This concerns, for example, the

choice of information which is considered relevant for a particular task, or what should be made explicit or left implicit in the message. Typically, these constraints are not of an obligatory nature; they simply propel the speaker in a certain direction when formulating a message, while still leaving a number of options.

One such option is the choice between presenting a given event as "on-going" or as "completed", illustrated above for English in (2a) and (2b), respectively. Traditionally, these two ways of presenting something are called "aspects", with the basic distinction between "imperfective aspect" and "perfective aspect". There is a vast and steadily increasing amount of research on the general notion of aspect and on these two aspects in particular (see Klein, this volume, section 3). Very roughly, there are three main lines of study that are interconnected in many ways. There are descriptive accounts of a number of aspectual systems, often treated in the context of the entire verbal system. If aspect is grammaticalized in a language, then any descriptive grammar of that language will say something about its form and meaning – as, for example, the many in-depth studies of the English progressive. Second, there are crosslinguistic and typological studies in which the expression of aspect is compared across a more or less extensive range of languages; the comparisons may be confined to a particular language, such as Slavic languages, where the differences at issue are usually quite subtle. There are also comparisons of languages that are typologically very different in which forms as well as meanings of aspectual markers diverge considerably, thus giving rise to the question of finding a common denominator for this category is (see, for example, Dahl 1985, 2000; Smith 1997). Third, there is a host of studies, many of them in formal semantics (see von Stechow, this volume), which address theoretical issues on aspect, or particular aspectual forms; one cannot claim that there is a shortage of theories on this topic (see, for example, Smith 1997 or Rothstein 2008).

In view of all of this work it is surprising how little agreement has been reached as soon as one goes beyond the most elementary level of description. This applies both to accounts of the empirical facts as well as the underlying theoretical notions. It holds even for languages that have been intensively studied in this regard, such as English or Russian. In both languages, aspectual distinctions are grammaticalized, and the morpho-syntactic facts of the distinction are well-described. In both languages, the differences in meaning between aspectually marked forms involve, in a broad sense, the familiar and well-established notions of imperfectivity and perfectivity:

    A. The situation is presented "from outside" versus the situation is presented "from inside".

    B. The situation is presented as "completed" versus the situation is presented as "non-completed" or "on-going".

    C. The situation is presented "with its boundaries" versus the situation is presented "without its boundaries".   (Klein, this volume, section 3)

But clearly, the Russian imperfective and the English progressive are not equivalent. Any translation of a Russian novel into English, or an English novel into Russian, makes this unmistakably clear; and any English speaker who has ever attempted to learn Russian, or any Russian speaker who has tried to learn English, is confronted with differences not only in meaning and but also in principles of use. If this is true for the imperfective in Russian and English, to what extent can we learn from descriptions of imperfective aspect in Chinese, for example, for which there are nevertheless quite a few empirical studies (see, for example, Xiao and McEnery 2004), compared to the many languages for which we have no more than one or two grammars written for practical purposes by a missionary or an isolated fieldworker? Aspectual characterizations may not be incorrect – but they are definitely not fine-grained enough.

    We believe that there are two main reasons for the present state of research in the field of aspect (and probably in the field of temporal expressions in general). They are both related to the way in which aspect is typically investigated. First, these investigations focus on the meaning of particular morpho-syntactic forms, for example the attachment of the (perfective) prefix *po-* in Russian or the *-ing* in English, and the combination of the resulting form with a copula, for example. With the focus on meaning and form, use of these forms is not analyzed in the context of the more complex processes given with the verbal description of specific events. This process is determined by a number of factors, in particular those mentioned above under A–D, as well as the interplay of these and perhaps other factors in determining the way in which a speaker "presents an event". Secondly,- and this point is related to the first – the traditional ways in which the relevant data are defined, as well as the way which the necessary empirical evidence is collected, restrict the empirical facts taken as the starting point for theoretical deliberations (see section 2 below).

    In this paper we present a method of analysis whereby established research procedures in the field of time semantics are complemented by experimental tasks. Speakers of different languages are asked to respond to an identical task (view a structured set of situations), with an input that is visual

and non-verbal, and thereby give a verbal account of what they have seen. This elicitation technique allows for the manipulation and analysis of the different steps involved between the perception of a particular visual stimulus on the one end and the utterance of a speech signal on the other. In the studies in question, speakers with different language backgrounds are asked to describe the same set of events. This event can be simple – a scene presented in a short video clip; but it can also be complex, such as the account of a car accident, which involves many sub-events that are temporally related to each other, with changes in places, objects, persons and the properties which these objects and persons have. Speakers are thus confronted with a verbal task in which all or many of the options outlined in (1)–(3), but also all or many of the constraints mentioned under A–D, come into play and can be investigated. It is also possible to study how speakers solve this task in real time, for example, by taking into consideration chronometrical parameters such as speech onset, or eye movements while viewing the scene, or in the course of language production, or both. Factors of this kind reflect the way in which attention is directed to the different components of the event. It is also possible to systematically analyze the final product – the sentence or sequence of sentences in which the event is described. This allows the investigation as to how patterns in event construal are adapted to the particular tools which a particular language offers. In other words, if there is a form of "language specificity", that is, a systematic influence given by the linguistic system in event representation, then this should become evident in the way in which speakers deal with one and the same task.

In illustrating the empirical approach we will focus on one aspectual distinction – the speaker's option of presenting an event as *on-going* or as not *on-going*. As mentioned above, this distinction is preliminary at this stage, and does not reflect the variety of actual aspectual meanings in different languages. We will report on findings for the Germanic languages English, German, and Dutch, as illustrations of the method used in investigating aspect. The methods can also be applied to the analysis of different temporal expressions, such as tense or temporal adverbials, and in particular to the interplay of these devices.

2. Some background considerations

As in any scientific endeavor, the study of how temporality is expressed is embedded in a given research tradition which shows an accumulation of an impressive body of knowledge. Any tradition may prove to be an obstacle

to a better understanding of the domain if it precludes the adoption of a fresh and unbiased look at the phenomena under investigation: we benefit, but we also suffer from tradition. It is thus useful to reflect briefly on the empirical and theoretical underpinnings of this tradition before taking up the thread once again – perhaps with a slightly different view.

Let us begin with the empirical side. What is the nature of the evidence on which the linguist's claims about the lexical and grammatical features of a human language are normally based? Essentially, there are two methods that are generally not strictly isolated but combined in different forms:

– looking at examples of language use, that is, written or spoken sentences or texts in a given language; this may be carried out in a systematic way by compiling and analyzing a data corpus; it may also be carried out on an ad hoc basis.
– appealing to the linguistic "intuitions" of someone who speaks the language (very often the linguist's own intuitions). This appeal can take the form of a grammaticality judgment (e.g., "Can you say that in Sorbian?") or a question about meaning (e.g., "What do you use this word for?").

Take, for example, a question such as: "What is the position of the finite verb in German, compared to Latin or Turkish?" A common way of answering this is to look at a more or less representative corpus of sample sentences. An initial inspection shows that the position of the verb in Latin is relatively free – it can appear at the beginning of the clause, at the end, and somewhere in-between. In Turkish, it is predominantly at the end – although other positions are found as well. In German, it also appears in various positions: *Hans stelle die Tasse auf den Tisch. – Stell die Tasse auf den Tisch! –…, weil Hans die Tasse auf den Tisch stellte.* Other possible positions are not observed. Both findings – the positions that are observed in the data, and the positions that are not observed in the data – are not fully satisfactory in answering the question. It could be accidental that certain positions are not observed (especially if the corpus of sample sentences is small). This can in part be remedied by changing the position and asking someone who speaks the language whether the resulting structure is possible – an appeal to the "intuitions" of this speaker. And as to the positions that are in fact observed – they yield a somewhat inconsistent picture. It is easy to see that the various positions are not random, but correlate somehow with structural and functional differences. But what are these differences? Here, the linguist resorts to his or her intuitions as someone who speaks the language. In this particular example, it is possible to correlate

varying positions with varying types of speech situations in which the utterance is used, and then relate the different positions to different functions.

An observational approach of this kind is hardly possible when it comes to the meaning of words, the meaning of bound morphemes such as a suffix, or the meaning of full constructions. In theory, one might try to determine the meaning of *to kill* or *to die* by observing all types of situations in which these words are used by someone (and in fact, this is essentially the way in which the meaning of these expression is learned). But first, this is practically impossible if more than a few words are to be described. Secondly, it is not very revealing, because one cannot easily determine which feature of the situation is related to that particular word. Finally, it is plainly impossible with expressions with a more functional meaning, be it morphemes such as *-ed* or *-ing*, or full words such as *the, still* or *or*. In practice, and sometimes in principle, there is so far no other option than to appeal to the intuitions of a competent speaker of the language in question; this appeal is supported by the clever use of specimens of the language when used in production. These can take the form of combinatorial tests. One might ask speakers (or oneself), for example, whether it is possible to combine a particular verb with the adverbial *for two hours*, and when these speakers' semantic intuition tell that this is not possible for *to explode* but possible for *to sleep*, we might take this fact as a basis for verb classification. This may be helpful, but only within certain limits. It cannot be readily transferred to other languages. And even in English, we have no satisfactory answer as to why such a combinatorial restriction applies and whether it is important (see Klein, this volume, section 4).

All of this is well-known. But if taken seriously, then it should not be surprising that our knowledge of how lexical aspect, grammatical aspect, as well as other temporal devices function in particular languages is still incomplete. Is there any way, any instrument, any measurement or procedure that would allow us to go substantially beyond what these two methods can provide us with? In fact, when it comes to meaning it is difficult to imagine how one could proceed without an appeal to the semantic intuitions of someone who "knows" the language. The meanings of *to work, beyond* or *-ing* are not visible, neither on paper nor in sound waves nor in the brain. But we can **refine** these two classical methods, in order to introduce a better level of control and comparability, and most importantly, to look at what speakers of some language actually do when describing a scene and use certain words and constructions to this end. This is the aim of the methods discussed in the following sections.

Let us now turn to the theoretical side. There are numerous, often very sophisticated and formally very elaborate theories of aspect. However, we are very far from a generally accepted theory of, say, the perfective-imperfective distinction. The aim of the following section is not to present another theory, but to point out some assumptions which theories have to consider. We will keep these assumptions to the minimum.

Consider 2 again, repeated here:

(2)   a. Brutus was killing Caesar with a dagger.
       b. Brutus killed Caesar.

These are two possible descriptions of one and the same situation. Intuitively, they have certain elements in common, and they differ in other respects. In both cases,

– there must be a person who does something, for example, uses a dagger; this person – the "agent", encoded here as subject – is called Brutus;
– there must be someone who is first alive and then dead; this person – the "patient", encoded here as the direct object – is called Caesar;
– the change of properties in the case of Caesar is caused by Brutus' activity;
– there is a inherent temporal structure; the time at which Caesar is alive precedes the time at which Caesar is dead, and the time at which Brutus does something must overlap with the time at which Caesar is alive; so, we have at least three inherent temporal intervals which are characterized in a particular way.

Note that the same inherent temporal structure is also found in sentences such as *Peter opened the door* or *Mary was painting the wall*. What differs are the entities that fill the argument positions, and the properties which go with the various intervals: doing something with the door handle, or the brush, rather than with a dagger; being first closed and then open, not green and then green; being alive and then dead. Thus, *to kill, to open, to paint* have the same argument structure, and they have the same inherent temporal structure, but the descriptive properties assigned to the arguments and the various temporal intervals which make up this structure are different. Other verbs, such as *to die, to cost, to watch, to give* may also have a different argument structure, and they may have a different temporal structure, and,

of course, different property assignments to the various arguments at the various time spans.

These considerations, simple as they are, lead to three important conclusions that must be observed in the study of temporal expressions:

1. Even with a simple verb, there can be an inherent complex temporal structure; in the examples just discussed, there are (at least) three temporal intervals which are related to each other by relations such as *later* or *overlapping*.

2. There is a clear difference between the *inherent temporal structure* given with an expression and the *descriptive properties* that go with the components of this structure, i.e., the various temporal intervals.

3. Temporal structure and descriptive properties of a lexical verb (and a more complex construction based on a verb) are relative to the arguments. The event encompasses different subintervals with different properties of the entities involved: it is Caesar who is dead at some subinterval of the event, whereas he was alive at some earlier subinterval of the same event.

Since all verbs do not require two arguments, they need not involve two time spans. In *Caesar died*, we have only one argument, but two time spans with the same descriptive properties, as specified for the second argument in *Brutus killed Caesar*: at the first time, he was alive, and at the second time, he was dead. The verbs *to kill* and *to die* are both what one might call "two state verbs"; but note that these two states – the source state and the target state or end state – are relative to specific arguments, here the second argument in *to kill* and the only argument in *to die*. Between these two states, there must be a transition, a point of transition, so to speak. The verb itself does not say anything about whether this transition is abrupt or smooth; so, the "transition point" need not really be point-like. In other verbs, the distinction between source state and end state is only gradual, as, for example, in *to rise*; if the temperature rises, it need not be "low" in the source state and "high" in the end state – it should only be "higher" in the second state than in the first. Note that this difference does not concern the temporal structure – in *to die* as well as in *to rise*, it involves a first time and a second time – but the descriptive properties which go with it: the difference between them can be categorical (be alive – be dead), it can also be gradual (lower – higher). The precise analysis of how this works with specific verbs is complicated and goes beyond the subject of this paper. But any analysis must somehow deal with these elementary observations.

Let us now return to the difference between *Brutus killed Caesar* and *Brutus was killing Caesar*. What has been said so far applies to both of these two (out of the many) options of describing one and the same event. Intuitively, the second version – the progressive – places us somehow in the midst of the action and the action is presented as "on-going". This does not preclude that it comes to an end; but the sentence *Brutus was killing Caesar* could be true without Caesar actually reaching the stage at which he is dead – the "end state" or "target state" of the entire action; someone could have intervened in due time. In *Brutus killed Caesar* this is not possible. The end-state or target state must indeed be reached, or the speaker's description is wrong. We can easily catch this intuition if we assume that the speaker's assertion is confined to a particular time – the time about which the speaker wants to talk, in this case, the time for which the assertion holds. In *Brutus was killing Caesar* this time must overlap with the time of Brutus' activity, but not the time of at which Caesar is dead – the end state of the second argument. In *Brutus killed Caesar* the time at issue must overlap with the time of Brutus' activity as well as the two other time spans – the time at which Caesar was alive (you cannot kill a dead man) and the time at which he was dead. In both cases the verb meaning of *to kill* includes an end state – a temporal of "Caesar be dead"; but in saying *Brutus was killing Caesar*, the speaker is not committed to the claim that this end state is reached; in *Brutus killed Caesar* he is committed to such a claim.[2]

Again, these observations are not particularly sophisticated. But they show that if we want to understand how aspectual distinctions function, we should not only look at the inherent temporal structure of verbs (and more complex verbal expressions), but also at the time about which the speaker makes the statement. Following Klein (1994) we shall call this time the "topic time". In the special case of assertions, we could also call it the "assertion time", i.e., the time to which the speaker wants to confine his statement. This choice is one of the many options which a speaker has when he or she sets out to describe a particular situation in a particular communicative context. It interacts with the lexical information provided with the verb – its inherent "argument-time structure" and the descriptive properties that go with it. It also interacts with the contribution made by other expressions to the entire meaning of a sentence, a point which we shall not take up here.

These background considerations serve to point to some of the facts which one has to keep in mind when investigating aspectual distinctions, as

---

2   Note that the sentence need not be true; the speaker could be wrong, or telling a lie.

mentioned earlier. In the next three sections we will illustrate how this can be carried out using empirical methods of analysis. We will show how speakers of three closely related languages, English, Dutch and German, proceed when describing motion events, as presented in video clips. We will then take a more differentiated look at other situation types and the extent to which they attract use of an aspectual viewpoint in these three languages.

3. Looking at the endpoint: the case of English, Dutch and German

Among the differences between the three main West Germanic languages, the degree to which aspect is grammaticalized is one of the most salient. In English there is a systematic distinction between the meaning conveyed by a progressive or continuous form, as in *John was closing the window, Peter was sleeping*, and forms which encode the completion of the event as in *John closed the window* or *John has closed the window*. With a few exceptions, this distinction applies to all verbs, and the speaker must choose between them when presenting an event: there is no neutral form (a good empirical survey is given in Williams 2002). German, on the other end, has no grammatical distinction of this kind. The simple forms *Hans schloß das Fenster, Peter schlief* can have both readings, and if he speakers wants to mark "on-goingness", then he or she has to use adverbials (such as *gerade* "just") or periphrastic constructions. The most important of these involve the prepositions *an* or *bei*, which both mean (roughly) "at". But the use of these forms is infrequent[3] and limited to some types of verbs: *Hans war dabei das Fenster zu schließen*, lit. "John was at-it the window to close" is fine, whereas *Peter war dabei zu schlafen, Peter war am Schlafen* or *Peter war beim Schlafen* are interpretable, but odd. Dutch is somewhere between these poles. Aspect is not grammaticalized, as in German, in the sense that its use is not obligatory in given contexts. The most common form to express the on-goingness of an event is the construction *aan het* + infinitive (*at the* + infinitive), which occurs with *zijn* (to be) as in *een man is viool aan het spelen* 'a man is violin <u>at the</u> play'. However, use of the associated aspectual perspective is considered to be much more common than in German (van Pottelberge 2004 is a comprehensive survey of the facts; see also Boogaart 1991; Booij 2002).

---

3   It is, however, considered typical for some German variants close to the Dutch border ("niederrheinische Verlaufsform", i.e., lower-Rhine progressive form).

This is no more than a rough sketch of the facts. It could be made more precise by (a) asking native speakers for their intuitions – both about what on-goingness means, whether they would use a particular form with a particular verb, or if they would rather use this or that form in a given situation, or (b) by looking at the use of these forms in actual corpora or spoken or written language. Both methods have their merits and their shortcomings (cf. section 2). In what follows, we will adopt a different approach – we will look at how English, Dutch and German speakers proceed when asked to describe one and the same set of events presented as on-going in a video-clip. The first set presented here involve situations showing *motion events* with a change in place such as *a person is walking down a street* where the end-state is a place (for example *a person walking to a house*). In traditional terms one could say that an event of this kind has a "right boundary". But the following points must be noted here. First, the "boundary" is not just the end of an event since is characterized by certain properties assigned to one of the arguments. These properties can be spatial (the entity is AT the house, whereas it was not before) – it is an end-place. The property could be of a more qualitative type (*the wall is painted*); a special case of the latter kind is that the referent of the argument only exists at the "target state": the house is now built, whereas it did not exist before. Furthermore, a difference must be made between whether the end-state is **mentioned** and thereby **marked as reached**. In English, for example, the description of one and the same scene could be:

(5)   a. Two nuns were walking.
       b. Two nuns were walking to a house.
       c. Two nuns walked to a house.

In (5b) as well as in (5c), the end-state – in this case the target position of the only argument, the nuns – is made explicit; but only in (5c), the time about which an assertion is made (the topic time) is large enough so as to include the end-state. If the end-state is not shown in the video as reached, the speaker need not necessarily zoom in on exactly what is shown – he or she could construe this phase as a part of a larger event, which includes the end-state. Then, the speaker has, so to speak, a holistic representation of the event: is is "viewed" it in its entirety, although it is not presented in the video in its entirety.

English, Dutch and German speakers (20 per group) were shown 60 video-clips and were asked to tell *what is happening?* They were also asked to start as soon as they recognized what was going on. 20 of the clips (test

items) showed situations involving a change in place of a figure (person or vehicle), while the rest were so called distractors. The speaker's utterances were tape-recorded and categorized by two independent observers as to whether the description included a reference to an "endpoint" of whatever type; this could include all sorts of end-states, spatial, qualitative, existence.[4]

3.1. Referring to the endpoint

The results for the three groups of speakers show a clear difference with respect to the verbalization of endpoints.

Table 1.   References to endpoints of motion events

English	Dutch	German
23.3%	15.5%	74.3%

English speakers tend to focus on the phase which is actually presented in the clip. If endpoints are not reached by the figure in motion, there is a clear preference not to mention them – although this is clearly possible, as in (5b) above. The –ing form is used in all cases (100%). In other words, since they were asked to tell what is happening, the event shown in the clips is perceived as on-going and the "zooming-in" on this phase goes so far that they do not mention a possible endpoint.

The results for German give a different picture. First of all, the event is not presented as on-going, although the system provides means to express this perspective. But constructions such as *am V-inf sein* or *beim V-inf sein* are not used to describe motion events, nor do they use adverbials to this end. They do not relate to on-goingness as a feature of events. Second, the show a clear tendency to refer to the endpoint, whether it is presented in the clip or just inferable. They do not "zoom into" a phase of the event but construe it as a whole; this holds even when it is left open as to whether the endpoint is actually reached within the time interval of the video clip. In

---

[4]   We have chosen to use the term "endpoint" in referring to the final point of a motion event since the term "(right) boundary" is used in different ways in the literature (see Klein, this volume). The term "end-state", on the other hand, is normally not associated with the final destination of a motion event.

other words, they take a holistic view on the events in contrast to the strategy of phasal decomposition in English.

Dutch differs from German in that on-goingness is indeed coded by *aan het V-inf zijn*. But in contrast to English, use of this perspective on-goingness is relatively low (see section 4 below). Restrictions in the use of on-goingness in Dutch can be explained by the fact that the relevant morphosyntactic construction is still evolving in the system. The grammaticalization of the periphrastic form *aan het* + infinitive in Dutch is reflected in the syntactic status of the prepositional phrase: in *een man is viool aan het spelen* the preposition *aan* is adjacent to the nominalized infinitive *het spelen*; the direct object *viool* precedes the entire prepositional phrase. This suggests that that the preposition is no longer head of a phrase which includes the object (see also Boogaart 1991; Booij 2002); it is "getting closer" to the verb, thus weakening its status as a locative preposition and approaching the status of an independent grammatical construction for on-goingness. Its semantic interpretation as a locative phrase governing a nominal phrase is bleaching into a verbal element; in this sense, it is becoming grammaticalized. The two constructions *viool aan het spelen zijn* and *viool spelen* are already quite parallel (note that in Dutch, as in German, the direct object precedes the non-finite verb). In German, by contrast, the prepositions have retained their locative status: *er ist dabei Violine zu spielen* "he is there-at violin to play" or *er ist beim Singen* "he is at-the singing"; one cannot say *er ist Violine am Spielen* or *er ist ein Lied beim Singen*.

Summing up, English, German and Dutch speakers show different preferences in the way the same events are conceptualized from a temporal perspective, in particular in the way they relate to the endpoint. The findings were investigated further with experiments testing non-verbal behaviour: chronometric analysis of speech onset times, as well as eye tracking while viewing the clips.

3.2. Comparison of speech onset times

This study is based on the same data. Speakers had been asked to start speaking as soon as they recognize what is going on in the video clip. The hypothesis was that if German speakers need an endpoint in order to satisfy the notion as to what can be coded as an event, they will wait for the event to become evident as a whole before starting to speak. By contrast, speakers of English, in which the aspectual distinction of on-goingness is coded

grammatically, can describe any phase of an event (inceptive, intermediate, terminative) and therefore do not have to wait for the endpoint an action.

Speech onset times for the test items were determined by measuring the onset of the sound wave in the digitalized version of these recordings.

Table 2. Speech onset times

English	Dutch	German
3.51 sec	4.0 sec	4.54 sec

On average, German speakers started speaking 4.54 seconds after stimulus onset (i.e. after the beginning of the video clip), while English speakers started about one second earlier ($t1(24) = 3.13$, $p = .004$; $t2(27) = 10.71$, $p < .001$). Dutch speakers are in between. This means that in order to get a reportable event, German speakers indeed wait longer, while English speakers are able to verbalize an on-going event such as *someone is walking* without having to figure out the endpoint (where the person is heading, for example). The results thus confirm what was found in the analyses of the verbal tasks: in providing the basis for a reportable event, German speakers show a clear preference for a holistic perspective, and this means waiting until the scene as a whole has unfolded before starting to speak. In contrast, any phase constitutes a reportable event for English speakers.

3.3.   Eye tracking

The specific hypothesis in this case is as follows: When verbalizing information on situations showing a figure on its way from one place to another, but where the goal in the video clip is not actually reached during the span observed, speakers of languages that do not have grammaticalized means to represent an event as on-going will not only refer to endpoints but will also be more likely to attend to them, compared to speakers of languages in which the temporal-aspectual concept "event is on-going" is grammaticalized.

Table 3. Number of fixations of endpoints before and after speech onset

	English	Dutch	German
Fixations before SO	2.9	4.06	6.9
Fixations after SO	8.5	5.59	9.5

The figures show a marked difference between German and English speakers in the number of fixations before speech onset, while Dutch is in between (see also von Stutterheim et al 2008, with results for seven languages). The German speakers require more information on the event before beginning to speak, since the event is construed in holistic terms with a possible endpoint.

Why is there such a pronounced difference between fixations before and after speech onset for the English group? English speakers apparently start speaking before they focus the endpoint and, as mentioned above, this can be linked to the fact that if motion events are viewed in aspectual terms, they are typically decomposed into phases (inceptive, intermediate, terminative phase). As with speech onset, the underlying phasal structure explains why English speakers – in contrast to speakers of German – do not tend to scan the scene for an endpoint from the outset. However, the eye tracking results show that in the course of the scanning process they visually control for possible end points. Speakers can add on the terminative phase and easily integrate it into the sentence which is already underway: *A car is going down a lane …… to a farmhouse*. These results provide evidence of a language-specific effect at the level of conceptualization, that is, before the speakers begin to speak.

4.   Presenting an event as on-going: Dutch and German

We have seen how speakers of English use an aspectual viewpoint when asked to tell "what is happening" and thereby respond to the phase focused in the video clip (intermediate phase). This means that they defocus the terminative phase, at first at least, and do not typically relate to the endpoint in the information verbalized. The questions pursued in the present section of the study are as follows:

– Are speakers of Dutch and German more likely to use an aspectual viewpoint in situations that do not have a prominent endpoint or transition point, as in situations listed under A below?
– How do they proceed if the situation has a prominent transition point with a target state, as in type B?

Type A Relatively homogeneous situations which last for a while and do not involve a salient change of any of the entities involved, e.g. a person surfing, jogging, or swimming;

Type B Situations which involve a salient "qualitative" change with respect to an entity (an effected object, as in building a monument, knitting a scarf; making a paper airplane); the scenes all show progression toward a qualitatively characterized target state;

Type C Motion events (going to the station) were again included in order to examine the extent to which they are viewed as on-going in both languages;

Type B has what one might call an "inner transition" with progression to a target state. This is given by the inferable final state of the object coming into existence, so to speak, whereas this is not the case for A. All three types are reflected in familiar Aktionsart classifications found in the literature (see Klein, this volume, section 4). It should be emphasized, however, that the classification into types, as given above, relates to situations shown in the video clips and not to distinctions which may be captured by linguistic expressions. So we are not talking about differences between verb types but situation types presented in video clips. This method allows us to systematically manipulate specific temporal properties of the events depicted so as to find out how speakers of the different languages respond to the same set of stimuli and the specific temporal features presented in them.

In contrast to English, where use of an aspectual perspective is 100% when viewing the same set of situations and telling what is happening, speakers of Dutch and German may represent the event as on-going, or not, as the following overview shows.

Table 4. Use of the aspectual viewpoint 'event is on-going' in Dutch and German

20 speakers	Use of an aspectual viewpoint
German	5.83%
Dutch	30.74%
English	100.00%

Dutch has other locative means, in addition to the periphrastic form, which include the verbs *zitten* (sit) and *staan* (stand), but use is relatively low in the present study (see also van Pottelberge 2004):

Table 5. Different aspectual constructions used – Dutch (% of total no. of responses)

	Aan het + V-inf zijn	Zitten/staan te + V-inf
L1 Dutch	237/911 – 26.02%	43/911 – 4.72%

We will now compare how one factor, the type of situation, in the sense outlined above, influences the choice made when taking a temporal perspective in Dutch and German.

Table 6. Viewing situation types as on-going in Dutch and German – situation types

20 speakers	Dutch	German
Type A (swimming)	(66/151) 43.71%	(20/151) 13.24%
Type B (moulding a vase)	(83/190) 43.68%	(8/190) 4.20%
Type C (going to the station)	zero	zero

English is again at 100% for all three situation types.

Type B situations were divided into groups with homogeneous sub-events (*knitting a scarf)*, and those with a higher range of heterogeneous subevents (*folding a paper airplane; baking a cake*), in order to test the role of homogeneity in adopting an aspectual viewpoint. In the case of knitting a scarf, the action shown in the video focused on the homogeneous movement of the hands. In the other group (such as baking a cake) different actions were shown such as stirring the cake mix and adding in flour, or folding the paper in different ways, straightening the wings, etc. The results reveal that B type situations with homogeneous sub-events have the highest rate in attracting use of an aspectual perspective.

Table 7. Type B Situations with homogeneous versus heterogeneous sub-events (5 videoclips each)

Aspectual perspective	Dutch	German
Type B Homogeneous subevents	48/76 63.16%	3/76 3.94%
Type B Heterogeneous subevents	17/76 22.37%	2/76 2.63%

The results for Dutch indicate that events showing a progression toward a resultant state (type B), and sub-events that can be viewed as homogeneous, are the most likely candidates in leading to the selection of an aspectual perspective.

Finally, the very few motion events (type C) that are represented aspectually as on-going in Dutch are confined to situations that can be viewed in a way similar to type A situations. The situation is represented without a directed change in getting from a specific point of departure to a possible goal. The event is viewed as "being out for a walk", with no mention of any goal, using a verb such as "wandelen". These occurrences were not included in the figures for goal oriented motion events (type C). It should be emphasised at this point that the mention of endpoints in motion events in the Dutch data may fluctuate depending on the extent to which speakers adopt an aspectual perspective in describing the set of scenes, as the entire series of other experiments on motion events have shown (see also von Stutterheim et al 2008 (submitted).

4.1. Discussion of the results

These findings give rise to a number of considerations. Dutch speakers are most likely to represent an event as going when the scene shows progression toward a qualitatively characterized target state – when a series of causative actions leads to the existence of a finished object through a progressive change in its qualitative properties (type B). Speakers also view the situation as on-going when the event is homogeneous and shows duration, without a progressive change, as in surfing or swimming situations (type A). Homogeneity is a core factor in type B as well as type A events, as the comparison between situations with homogeneous versus heterogeneous events (type B) show.

Although figures are still very low for German, there is emerging evidence that type A situations also attract use of an aspectual perspective, but not when the situation shows a progressive property leading to a resultant state. The findings thus indicate that the aspectual distinction encoded in Dutch encompasses what may be called "progressive" as well as "on-going" components, while the relevant criterion in German may centre on on-going-ness – as represented by homogeneous events with a clear duration and no qualitative change.

Finally, the very few motion events (type C) that are represented aspectually as on-going in Dutch are confined to situations that can be viewed in

a way similar to type A situations. The situation is represented without a progressive change in getting from a specific point of departure to a possible goal: it is viewed as "being out for a walk" with no specific goal using verbs such as "wandelen". Representation in this way conforms with a possible constraint that calls for homogeneity and duration, but not changes leading to a resultant state. So why is the form of progression (change in place from source to goal) given with motion events not yet accessible for use with an aspectual perspective, in contrast to causative (type B) events? The answer may lie in the nature of the transition. In contrast to motion events, causative actions, as presented in type B situations, may constitute a prototypical context for the concept of progression, since they provide a tangible contrast between the pre-state of the event and its progression to the final post state.

The results presented here for Dutch and German form part of a larger study on aspectual concepts in Romance (Natale 2008; Carroll et al. 2008; Leclerq 2008) as well as Semitic (Arabic) and Slavic languages (v. Stutterheim et al submitted).

5. Concluding remarks

Aspect is an important, but also a very difficult temporal category, and studies on how it should be defined and on how it is realized in different languages are legion. But we are far from reaching agreement on what is involved, except on a very global level, and our knowledge about the form and function of aspects in particular linguistic systems is far from satisfactory. Statements such as "language x is an aspect language" or "language y has an imperfective aspect", may not be false, but they hide more problems than they answer. The aim of this chapter was to present an empirical approach to the study of aspectual phenomena. Rather than depending on the semantic intuitions of someone who knows the language, or examinations of the use of aspectual forms in corpora, empirical analyses were conducted on how speakers proceed when solving different verbal tasks. Tasks can be systematically modified in many ways, and with this the range of analysis of the resulting data. The few findings presented here indicate how a procedure of this kind may lead to a more refined picture of aspectual distinctions, compared to traditional methods. It can not replace these methods, the appeal to intuition and the examination of corpora remains indispensable, but they have their limits and research procedures should try to go beyond them.

References

Boogaart, Ronny
    1991   "Progressive aspect" in Dutch. In *Linguistics in the Netherlands,* Frank Drijkoningen and Ans van Kemenade (eds.), 1–9. Amsterdam: Benjamins.

Booij, Geert
    2002   Constructional Idioms, Morphology, and the Dutch Lexicon. *Journal of Germanic Linguistics* 14: 301–329.

Carroll, Mary, Natale, Silvia and Starren, Marianne
    2008   Acquisition du marquage du progressif par des apprenants germanophones de l'italien et néerlandophones du français. *AILE*, 31–50.

Dahl, Östen
    1985   *Tense and Aspect Systems*. Oxford: Blackwell.

Dahl, Östen (ed.)
    2000   *Tense and aspect in the languages of Europe.* Berlin/New York: Mouton de Gruyter.

Drijkoningen, Frank and Ans van Kemenade (eds.)
    1991   *Linguistics in the Netherlands.* Amsterdam: Benjamins.

Klein, Wolfgang
    1994   *Time in Language.* London: Routledge.

Leclerq, Pascale
    2008   L'influence de la langue maternelle chez les apprenants adultes quasi-bilingues dans une tâche contrainte de verbalisation. *AILE*, 51–70.

Natale, Silvia
    2008   Semantische Gebrauchsdeterminanten der progressiven Verbalperiphrase stare+gerundio. Eine datenbasierte Studie. Dissertation, University of Heidelberg, Germany.

van Pottelberge, Jeroen
    2004   *Der am-progressiv*. Tübingen: Narr.

Rothstein, Susan (ed.)
    2008   *Theoretical and Crosslinguistic Approaches to the Semantics of Aspect*. Amsterdam: Benjamins.

Smith, Carlota
    1997   *The Parameter of Aspect*. Dordrecht: Kluwer. 2nd edition.

von Stutterheim, Christiane
    1997   *Einige Prinzipien des Textaufbaus*. Tübingen: Niemeyer.

von Stutterheim, Christiane and Carroll, Mary
    2006   The impact of grammatical temporal categories on ultimate attainment in L2 learning. In *Educating for Advanced Foreign Language Capacities*, H. Byrnes, H. Weger-Guntharp and K. Sprang (eds.), 40–53. Georgetown: Georgetown University Press.

von Stutterheim, Christiane, Dorothea Bastin, Mary Carroll, Monique Flecken and
Barbara Schmiedtová
    2008   How grammaticized concepts shape event conceptualization in the
                early phases of language production: Insights from linguistic analysis,
                eye tracking data and memory performance. Submitted to *Cognition*.
Williams, Christopher
    2002   *Non-progressive and progressive aspect in English*. Bari: Schena.
Xiao, Richard and Tony McEnery
    2004   *Aspect in Mandarin Chinese*. Amsterdam: Benjamins.