*Ronald R. Jacobsen**

The role of knowledge systems in the linguistic construction of action scenes in novels and their translations

Abstract

What makes one translation better than another? This paper argues that the best is the one that best mirrors the levels of organisation found in the source text while at the same time achieving coherence on as many of them as possible. For instance, in the English translation of Peter Hoeg's novel *Smilla's Sense of Snow*, the original Danish sentence which contains a BE-perfect, *Med Esajas i sin kiste er kommet et følge* ['With Isaiah in his coffin is come a procession'] (Høg 1993: 11) is translated as: *A procession follows Isaiah in his coffin* (Hoeg 1993: 4). This translation may achieve grammatical and local coherence, but certainly not global coherence since it involves a re-construal of the preceding text as a dynamic, or evolving, scene thereby clashing with the static one constructed in the source text. Likewise, it disagrees with the propositional content of the source text on several levels. This may be as it is, but the real problem is: how do you model the relations between the several levels of organisation found in a text like this in order to qualify/support a particular translation? The answer to this problem, this paper argues, is mental space (MS) theory. Accordingly, an outline of a very detailed analysis of the action scene constructed in the beginning of Don DeLillo's novel *Underworld* (DeLillo 1999a) is presented and compared with its Danish translation (DeLillo 1999b).

1. Introduction

This paper presents an analysis of the action scene constructed in the beginning of Don DeLillo's novel DeLillo (1999a: 11-14) and its Danish translation DeLillo (1999b: 7-10). Its aim is to demonstrate not only how the interaction between the four main knowledge systems participating in the construction of the action scene but also between

* *Ronald R. Jacobsen*
The Aarhus School of Business
Fuglesangs Allé 4
DK-8210 Aarhus V
rrj@asb.dk

the different levels of organisation involved can be modelled in terms of mental space (MS) theory, cf. Fauconnier (1994, 1997) & Jacobsen (2004). The four main systems are: (1) one's basic experience with three-dimensional (3D) space; (2) one's experience with the linguistic system of modern American English, in particular its tense/aspect system; (3) one's experience with contemporary Western film and its conventions, for instance, how action scenes are represented visually; and (4) one's experience with literature and its conventions, for instance, how to recognise Free Indirect Discourse (FID), narrative voice, etc.. Subsequently, it is shown how this modelling of the source text can be used as a basis for judging the quality of a translation. That is, the best translation is said to be the one that best mirrors the levels of organisation found in the source text while at the same time achieving coherence on as many of them as possible.

The source text can be described as follows. Firstly, it consists of seventeen paragraphs the transition between each of which involves a change in perspective. Secondly, it concerns the gate-crashing of fifteen boys who want to enter a stadium in order to watch a base ball game. The protagonist is the youngest of them, Cotter Martin, who is a fourteen-year old black kid from Harlem. Thirdly, the most salient property of the text is the complex interaction it triggers between the perspectives established directly for the three-dimensional (3D) story world and the narrative voices the reader is prompted to construct, as indicated by the following brief description of each of the seventeeen paragraphs of the text:

Paragraph	**MNT**[1]	**Content**	**Narrative Voice(s)**
1		Description of the protagonist ("what's in the frame")	Narrator's local perspective
2	Yes	Description of the local setting ("looking around")	Narrator's local perspective
3	No	Description of the global setting (timeless or universal perspective)	Omniscient narrator's perspective

[1] "MNT" is the acronym of Movement of Narrative Time, and the labels "yes" and "no" describe whether the transition from one paragraph to the next involves a movement in narrative time or not.

4	Yes	Description of the local setting (“looking up”)	Omniscient narrator’s perspective
5	Yes	Description of the protagonist and his accomplices (“looking around”) + mention of a projected course of events	Omniscient narrator’s perspective
6	Yes	Description of the local setting (prelude to the main action)(expanded “looking around”)	Omniscient narrator’s perspective
7	Yes	Description of the local setting (initiation of the main action)(“looking from outside the scene”)	Omniscient narrator’s perspective (“objective” account of what happens)
8	Yes	Description of the local action-scene and the protagonist, stage 1 (viewpoint = Cotter’s)	Omniscient narrator’s perspective (protagonist’s perspective = the narrator’s perspective)
9	Yes	Description of the local action-scene, stage 2 (viewpoint = Cotter’s visual input)	Ambiguous between an embedded narrator’s (the protagonist’s) perspective and an omniscient narrator’s perspective
10	Yes	Description of the local action-scene, stage 3 (viewpoint = Cotter’s auditory input)	Ambiguous between an embedded narrator’s (the protagonist’s) perspective and an omniscient narrator’s perspective
11	Yes	Description of the protagonist’s action (preparation for main action)	Embedded narrator’s perspective followed by omniscient narrator’s
12	Yes	Description of the protagonist’s action + mention of projected action (i.e mental simulation of the main action)	Ambiguous between an embedded narrator’s (the protagonist’s) perspective and an omniscient narrator’s perspective
13	Yes	Description of the protagonist’s main action, stage 1	Embedded narrator’s perspective followed by omniscient narrator’s
14	Yes	Description of the protagonist’s main action, stage 2 (climax)	Omniscient narrator’s perspective (who is mimicking the protagonist)

15	Yes	Description of the protagonist's emotional response	Omniscient narrator's perspective (who is mimicking the protagonist)
16	Yes	Description of the protagonist's action and the new local setting inside the stadium (postlude)	Omniscient narrator's perspective
17	Yes	Description of the protagonist's action and the local setting (fade-out)	Omniscient narrator's perspective

2. The interaction between the four main knowledge systems

Mental space theory is founded on two basic assumptions. The first, from which the name of the theory derives, is that mental representations are structured into interconnected mental spaces. Put less formally, mental representations which are constructed for local purposes as we speak and think are assumed to be divided into smaller conceptual packets. Mental spaces are therefore defined as: "structured, incremental sets—that is, sets with elements (a, b, c, …) and relations holding between them (R1ab, R2a, R3cbf, …) such that new elements can be added to them and new relations established between their elements" (Fauconnier, 1994: 16). The second assumption is that language only provides cues or instructions for such mental space (or meaning) constructions. A major advantage of this theoretical framework is that it allows one to study/model not only how on-line meaning construction relies on knowledge already stored in memory, i.e. on existing semantic frames (read: knowledge systems), but also how this knowledge comes into existence in the first place.

Evidently, in order to capture the interaction between the four knowledge systems, one needs some model of their interface—i.e. of how they manage to "communicate". The applications of MS theory, including blending theory, evidenced so far do not adhere to any particular psychological theory of knowledge. Hence, this paper proposes that Lawrence W. Barsalou's (1999) theory, *Perceptual Symbol Systems*, be adopted[2]. This theory assumes that "cognition is inherently percept-

[2] See also Glenberg (1997) and Goldstone & Barsalou (1998).

ual, sharing systems with perception at both the cognitive and the neural levels" (Barsalou 1999: 577). A concept is said to be equivalent to a **simulator**, and a conceptualisation to a **simulation**. For instance, Barsalou (1999: 587) says:

> "According to this theory, the primary goal of human learning is to establish simulators. During childhood, the cognitive system expends much of its resources developing simulators for important types of entities and events. Once individuals can simulate a kind of thing to a culturally acceptable degree, they have an adequate understanding of it. What is deemed a culturally competent grasp of a category may vary, but in general it can be viewed as the ability to simulate the range of multimodal experiences common to the majority of a culture's members. Thus, people have a culturally acceptable simulator for *chair* if they can construct multimodal simulations of the chairs typically encountered in their culture, as well as the activities associated with them."

In other words, the interface between the different knowledge systems is explained in terms of a common representational system of perceptual symbols. Given this assumption, the four main knowledge systems and their interface can be explained thus: (1) one's basic experience with 3D space establishes an overlapping perceptual and conceptual system. For instance, when experiencing a situation in the world directly, a simulation is run which is mapped onto the sensory input thereby influencing how the situation is perceived. Likewise, a simulation can be run in the absence of sensory input, as in planning or imagination. This system is pre-linguistic. (2) One's experience with the linguistic system of modern American English develops simulators for words which are linked to simulators for the entities and events to which they refer (or other aspects of simulations). Once established, they can control simulations. But, there is also another fundamental way in which language is linked to the first, or basic, knowledge system. In contrast to fishes and frogs and some mammals who have their eyes on the sides of their heads, humans have their eyes in the front of their heads; which means they have to turn their heads to get a full view of the surroundings (Anderson 1996: 91). Therefore, humans have a capacity for integrating a succession of partial views into a world that is <u>continuous</u> in both space and time. This capacity, it is argued, has certain implications not only for the human perceptual symbol system but also, as a consequence thereof, language processing. Hence, Jacobsen (2004)

proposes: (a) the **Continuity Principle**, which states that if a situation[3] *a* is linked to another situation *b* by a connector *F*(*a*, *b*), then *a* & *b* are part of a complex situation *c* and a description of either *a*, *b* or *c* may therefore be used to identify one of the other two situations; and (b) three primary classes or types of continuity involving eight basic **pragmatic functions**, or **connectors** which link the situations identified by the sentences of a text, i.e. they describe what kind of continuity is selected (and as a consequence which representation is constructed). The first class concerns MOTION, and it includes: (i) the object connector, F(o), (ii) the action connector, F(a) and (iii) the double-action connector, F(da). The second class concerns ORIENTATION, and it includes: (iv) the inter-object connector, F(io); (v) the inter-action connector, F(ia); (vi) the object-inter-action connector, F(oia). The third and final concerns HIERARCHY, and it includes: (vii) the spatial-hierarchy connector, F(sh); and (viii) the temporal-hierarchy connector, F(th). For instance, the connectors explain why the following two conjoined sentences, *I took out my key and opened the door*, can be understood as either one continuous action, F(a); two independent but sequentially ordered actions, F(da); two temporally overlapping actions, F(sh); or a habitual event, F(th). Naturally, the constraints imposed by the linguistic system on which connector can be selected differ from one language to another. (3) One's experience with contemporary Western film and its conventions develops simulators for (far) more complex sequences of events, i.e. action scenes, than those mentioned in (1). As Joseph Anderson (1996: 112-113) points out:

> "A great deal of research has been done on proprioception and orientation, and the major outcome, at least for our purpose, is that invariably when one of these systems is in conflict with vision, it is vision that dominates. Visual motion can make us seasick, visual distortions of verticality and horizontality can make us stumble, and the appropriate visual information in a movie can make us feel that we are floating in space. Only because vision dominates proprioception is it possible for a motion picture to provide the information for self-movement. **Only because the visual array provides compelling information to the visual system for changes of viewer position is it possible for us to integrate a succession of shots from several**

[3] Accordingly, henceforth, the term *situation* shall be taken to concern the mental representation of something, not of some thing.

> **different camera positions into one unified visual inspection of the scene**."

This makes the third system an expert system since without an appropriate level of visual training, it is difficult to imagine how the complex simulations[4] under consideration could be produced, or computed, with ease in the absence of the stimuli [see (1)]. In fact, the relevance of which shall become apparent later, this paper would argue that even with extensive training, this sort of computation is very difficult in the absence of visual stimuli. So far, only the interface between systems 1 and 3 has been explained. The interface between 2 and 3 is established by the fourth system. (4) One's experience with literature and its conventions develops simulators for larger textual structures. Hence, system 4 could be argued to be an extension of system 2. This makes it an expert system. That is, without the appropriate level of training, it is difficult to imagine how a literary text involving, for example, FID, different narrative voices, etc., could be processed with great ease—since the textual simulators required would be missing.

This leads to the action scene under consideration. It differs significantly from the typical ones found in novels. That is, typical "action" scenes are mainly concerned with either the events leading up to the action, with what happens inside the protagonist's head or a combination of these, and they only "sum up" the main temporal and spatial coordinates of the action involved. Consider, for instance, the following classical example of "summing up" from the Old Testament (I Samuel, verse 17: 48-49), David's slaying of Goliath:

> (1) The Philistine drew steadily closer to David to attack him, while David quickly ran toward the battle front to attack the Philistine. David reached his hand into the bag and took out a stone. He slung it, striking the Philistine on the forehead. The stone sank deeply into his forehead, and he fell down with his face to the ground[5].

That is, 47 verses of background description and dialogue precede this description. Therefore, if one takes into consideration the rare occurrence of "real" action scenes in novels, it seems that even experienced readers

---

[4] By a *complex simulator* is meant a simulator which links or combines the simulators of individual concepts.

[5] English translation downloaded from Netbible: http://www.bible.org/cgi-bin/netbible.pl

are bound to face difficulty in processing such a text for the simple reason that they lack the required textual simulators. Notice that this explanation only serves as a motivation for the analysis proposed. Naturally, the rare occurrence of "real" action scenes in novels may be a consequence of the cognitive difficulty associated with running this type of simulation in the absence of visual stimuli, as already mentioned above. Hence, it is argued that the difficulty associated with reading the text has to do with identifying textual structures which can be aligned with a particular filmic **frame**[6], or schema, of action activated in memory (recall that high-level perception is consider equivalent to cognition by *Perceptual Symbol Systems* (Barsalou 1999)). Notice that according to Barsalou (1999), the cognitive processes underlying this alignment need not be conscious since what is referred to is only neural activity. It follows that the linguistic simulators linked to one's basic experience with 3D space are deemed insufficient for comprehending the action scene analysed[7]. In addition, as shall be demonstrated in the next section, textual simulators, like those associated with the recognition of FID and narrative voice, play an important role in recruiting from memory the filmic frames required in the alignment process mentioned above.

3. Levels of organisation

The highest level of organisation is the generic story frame which is shared by both film, system 3, and narrative, system 4. They differ mainly in the way the generic frame is realised. Hence, it seems fair to assume that both systems possess sub-frames which fill in the generic frame in on-line meaning construction. But there are also some striking

---

6 From its conception, MS theory has adopted a rather broad definition of *frame* which includes a number of different approaches to frames, scripts, schemata etc. (Fauconnier 1994: 10). However, despite being used in the same "loose" way in this paper, it's use will nevertheless be assumed to be in agreement with Barsalou's (1999) understanding of the term in the analyses presented below. Barsalou (1999: 590) offers the following broad definition of the term: "A frame is an integrated system of perceptual symbols that is used to construct specific simulations of a category". In other words, a frame is a *complex simulator*, see footnote 4.

7 This said, other types of expert knowledge, such as, e.g., experience with the different participant roles associated with a football game, may very well provide simulators which allow an easy processing of the text.

parallels between the sub-frames of these two systems, a point which is elaborated below. The lowest level of organisation is the sentence or clause and its elements.

As regards film, the relevant frames with respect to the problem at hand, the construction of an action scene, concern the integration of successive shots: what film theoreticians call (continuity) editing. Bordwell & Thompson (2001: 251) list four dimensions of film editing: (1) **graphic relations** between shot A and B, (2) **rhythmic relations** between shot A and B, (3) **spatial relations** between shot A and B, and (4) **temporal relations** between shot A and B. The first concerns the purely pictorial qualities of the two shots (lighting, colour, setting, costume etc.). The second concerns the length of each shot. For instance, Bordwell & Thompson (2001: 257) say: "each shot, being a strip of film, is of a certain length, measured in frames, feet or meter. And the shot's physical length corresponds to a measurable duration onscreen". Needless to say, rhythm in cinema includes music (as well as other factors) and this may influence rhythmic editing. The third concerns the spatio-temporal location of the situations represented by each shot with respect to each other. The following types of cut belong to this category: crosscutting, shot-reverse-shot, over-the-shoulder shots, eye-line matches and Point-Of-View (POV) shots[8]. The fourth, and final, concerns order, duration and frequency (Bordwell & Thompson 2001: 260). 'Order' includes sequences of shots that represent events in their actual order of occurrence as well as shots that establish flashbacks and flashforwards. 'Duration' concerns the length of story events, for instance, a part of a character's action may be cut away. This is called 'temporal ellipsis'. 'Frequency' includes repetition of a shot or a part of it, for example, a phase of an action. Clearly, language comprehension is not governed by any graphic concerns, but otherwise the same kinds of relations as those referred to by the terms 'rhythmic', 'spatial' and 'temporal' are part of it. Thus, the following linguistic choices of the text may be said to establish its rhythm: (1) tense(/aspect), (2) syntax and (3) lexemes referring to sound and dynamic events. The first choice mainly concerns the fact that the text is a narrative in the historical present (Notice there is overlap between systems 2 and 4 here). This structure is strongly

[8] The reader is encouraged to consult Bordwell & Thompson (2001) for descriptions of these different types of cuts.

associated with "eye-witness accounts", such as personal experiences, sports commentaries, stage directions and jokes. In addition, a narrative in the historical present gives the reader the illusion of being in the middle of the event described. How could this have anything to do with rhythm? The second kind of linguistic choice answers this question: in being associated with a kind of eye-witness account, the syntactic 'style' of the text, i.e. mainly the syntactic complexity and 'length' of the individual clauses, may not only mimic spoken language but also, as a result, the narrator's psychological involvement. In other words, long complex sentences are found in paragraphs mirroring shots of a certain length (a slow pace) and short simple sentences are found in paragraphs mirroring a sequence of briefer shots (a fast or accelerating pace). Like in language comprehension, the remaining types of cuts, i.e. those Bordwell & Thompson (2001) classify as either concerning spatial or temporal relations, can all be explained in terms of one or several of the eight connectors mentioned in section 2.

As regards literature, a number of frames have already been mentioned, i.e. that of the historical present, the kinds of continuity underlying a 'sequence of events', either F(a) or F(da); 'inter-action', F(ia); 'background description', for instance, of the scenery, F(sh); but an important frame remains: the one associated with the identification of narrative voice which resembles the POV shot in that it prompts for the same cognitive operations—in MS terms both set up a viewpoint space (Cutrer 1994: 73-74). The reason it is important is that it has a heavy influence on how the action is visualised[9], and hence also which filmic frame(s) may, or may not, be activated in the reader's mind.

Compare the following three paragraphs from DeLillo (1999: 11, 12, 13):

(2) It's a school day, sure, but he's nowhere near the classroom. He wants to be here instead, standing in the shadow of this old rust-hulk of a structure, and it's hard to blame him—this metropolis of steel and concrete and flaky paint and cropped grass and enormous Chesterfield packs aslant on the scoreboards, a couple of cigarettes jutting from each.

[9] Notice that by visualisation is not meant a picture corresponding to that produced by film but rather the neural activity underlying a simulation run by a simulator which has been established as a consequence of watching films.

(3) They are at the curbstone, waiting. Their eyes are going grim, sending out less light. Somebody takes his hands out of his pocket. They are waiting and then they go, one of them goes, a mick who shouts Geronimo.

(4) Then he leaves his feet and is in the air, feeling sleek and unmussed and sort of businesslike, flying in from Kansas City with a briefcase full of bank drafts. His head is tucked, his left leg is clearing the bars. (…).

Paragraphs (2) and (3) show the difference between the linguistic representation of a situation corresponding to a slow shot and one corresponding to a fast, or accelerating, shot. They differ in: (1) sentence length, (2) lexical references to sound and/or dynamic events, e.g. *He wants to be here instead* versus *then they go*, *shouts Geronimo*, and (3) kind of continuity selected, i.e. object, spatial-hierachy and temporal-hierarchy connections versus object, action and double-action connections, respectively. The third paragraph, (4), shows the filmic potential of narrative voice in that *flying in from Kansas City with a briefcase full of bank drafts* can be constructed in two ways. That is, it could either be constructed as the protagonist's "feeling", i.e. be invoked in a thought space, or as a (metaphorical) description, or elaboration, of 'being in the air'. In the first case, it is identified as the protagonist's voice, in the second, as the narrator's. In a corresponding filmic representation of this non-finite adverbial clause, one would either choose to leave it out or to represent it by a POV shot if the first interpretation is selected and one would choose to represent it by what is called a 'nondiegetic insert' if the second interpretation is selected (Bordwell & Thompson 2001: 281-284)[10]. This ends the discussion of the cognitive alignment of the filmic and literary frames involved in the consumption of the text. The next section reports the results of an investigation of a Danish translation of *Underworld* (DeLillo 1999a: 11-14), DeLillo (1999b: 7-10), which focuses on the clash between the English and Danish tense/aspect systems.

[10] A **POV shot** is defined as *A shot taken with the camera placed approximately where the character's eyes would be, showing what the character would see; usually cut in before or after a shot of the character looking* (Bordwell & Thompson 2001: 433). A **non-diegetic insert** is defined as *A shot or series of shot cut into a sequence, showing objects represented as being outside the space of the narrative* (Bordwell & Thompson 2001: 432).

4. The Danish translation

In section 2, the interface between the four main knowledge systems participating in the text's reception/comprehension was described in terms of a shared representational system of perceptual-symbol-systems' simulators.

System 1 was said to comprise the simulators established by one's basic experiences with 3D space. For instance, one's pre-linguistic concept of RUN could be argued to involve a complex simulator, **a frame**, consisting of two simple simulators, the first derived from one's basic visual experience of somebody else running and the second derived from one's own, or basic proprioceptive, experience of running. Naturally, a simulation run in the presence of the above-mentioned stimuli differs from a simulation run in its absence. The important point to notice, however, is that both types of simulation are assumed to share the same basic computational structure. Admittedly, a more realistic characterisation of one's concept of RUN would involve other basic experiences, including culture-specific ones, apart from those describable in terms of a 3D algorithm. For instance, to a Danish child the concept may be strongly associated with playing a ball game, such as soccer or handball, whereas to a Zambian child living in the bush the concept may be strongly associated with going to school or hunting down a domestic animal (read: cultural frames). In other words, the basic frame associated with the concept of RUN would involve a more complex simulator than the simple two-component's simulator described above. However, no matter how the simulation is run, in the presence or in the absence of stimuli, or how complex the frame associated with RUN turns out to be, the following general characterisation offers itself: the concept involves an indefinite number of situations linked by action connectors (see figure 1 below). This reveals two important aspects of MS theory: (1) pragmatic connectors are only part of semantic frames, that is, they describe generalisations across contexts, and (2) since experiences (read: semantic frames) vary from community to community, so may the pragmatic connectors available with a particular "concept" (Fauconnier 1994: 10).

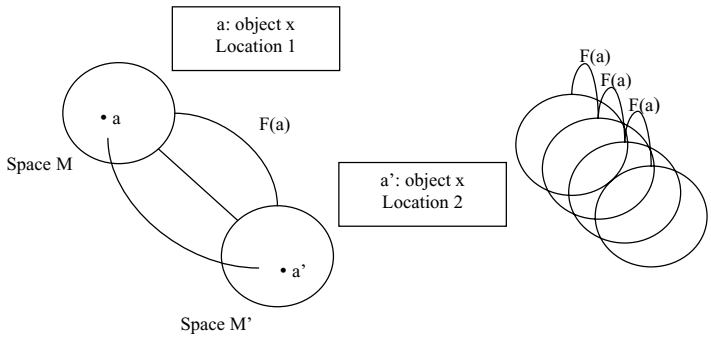

Figure 1. The concept RUN. The two spaces on the left connected by a single action-connector, F(a), illustrate one instance of the indefinite number of situations connected by action-connectors which underlie the concept of RUN as shown by the four spaces on the right. The element *a* in space M and its counterpart *a'* in space M' indicate the individual who runs. In space M, *a* enters one relation, identified in the upper box as location 1, and in space M', *a'* enters another relation, identified in the lower box as location 2, i.e. the individual identified as 'a' has moved.

System 2 was said to comprise the simulators established by one's experience with the linguistic system of modern American English. In contrast to system 1, the stimuli, i.e. the linguistic signals, activating system 2 simulators do not relate directly to the situations experienced. For instance, upon hearing the word *run*, one does not have a primary experience of the phonetic string /r∧n/, but rather of the simulation run by the system 1 simulator for RUN mentioned above. This implies that system 2 simulators by definition are complex simulators. In other words, upon hearing the word *run*, the simulator associated with the phonetic string /r∧n/ is activated and a simulation of /r∧n/ is run which is then matched with the sound received by the ear. Thereupon the simulator associated with the concept of RUN is activated and a simulation of RUNNING is run. An important parallel between systems 1 and 2 is that the Continuity Principle operates at both levels of cognitive organisation. That is, not only the situations constructed in visual cognition are connected by pragmatic functions, see figure 1, but also the situations constructed in auditory cognition, i.e. sentences or utterances. In this context, it is worth noticing that without such connections, the instructions offered by tense/aspect markers in

language would have no informational value. Consider, for instance, the simple past tense in English. It presupposes two situations which are temporally related, the one prior to the other, i.e. a temporal-hierarchy connection. Consider also the progressive aspect in English. It presupposes a situation embedded in another situation. That is, it either requires the establishment of a spatial-hierarchy connection or a temporal-hierarchy connection.

System 3 was said to comprise the simulators established by one's experience with contemporary Western film. In contrast to system 1, the stimuli, i.e. the strips of film (shots) spliced together, activating system 3 simulators do not convey situations which can be experienced by any single individual in a natural environment. For instance, in a typical Hollywood production, an action scene even of a limited duration contains a very high number of cuts involving not only many shifts in camera position/viewpoint but also slow-motion shots as well as shots/cuts constructed by other techniques, cf. the quotation by Anderson (1996: 112-113) in section 2. In other words, system 3 includes a type of complex 3D simulators, or frames, not firmly established by system 1. Like systems 1 and 2, the situation descriptions, i.e. shots, controlled by this system are governed by the Continuity Principle and its associated pragmatic connectors.

System 4 was said to comprise the simulators established by one's experience with literature. In order to distinguish these from the **linguistic simulators** covered by system 2, one might choose to call them **textual simulators**. Especially, the central role played by FID and voice was stressed in sections 2 and 3.

In section 3, three levels of organisation were distinguished, from higher to lower: (1) the filmic, (2) the literary and (3) the linguistic (or sentence level). That is, the filmic is assumed to dominate the interpretation of the literary and linguistic levels. This is consistent with the premise that the intended reader/recepient of the text is somebody who knows expert systems 3 and 4 and that the main source of entertainment is the text's potential for filmic interpretation (i.e. as a simulation of an action scene), which is the point of departure for judging the quality of the translation. In a typical translation study, the main concern would probably be the incompatibility between a, or several of the, non-linguistic knowledge system(s) of the source and target culture.

For instance, if one had studied the translation of DeLillo (1999a) into some Indian language, the main concern would probably have been the difference between Hollywood and Bollywood (read: Bombey) film. However, since there are strong historical ties between Anglo-American and Danish culture(/language) and since the kind of film produced and consumed in the USA and Denmark closely resemble each other, the focus can be shifted from the non-linguistic knowledge systems, i.e. systems 1, 3 and 4, to the linguistic system, i.e. system 2. This leads to the Danish tense/aspect system.

Four features of the source and target texts which typically would be treated as relating to tense and aspect are: (1) the use of the historical present, (2) the use of perfect "aspect", (3) the use of progressive aspect and (4) the role played by lexical aspect. Since a grammatical morpheme as a general principle is associated with a single instruction or set of instructions by the theoretical framework outlined in this paper, there is no room for treating the historical present as a separate tense. Instead, as already mentioned in section 3, it is considered a literary frame or **viewing arrangement**[11] which gives the reader the illusion that he or she is in the middle of the event described. As regards the second feature of the texts, the use of perfect "aspect", it is not considered relevant to the issue at hand for two reasons. Firstly, this paper agrees with Mourelatos (1981) that the meaning expressed by the form does not primarily concern aspect but phase. Secondly, even though Danish has two perfect forms, a HAVE perfect and a BE perfect, it seems that the two forms only pose a problem when translating from Danish into English, especially if the translator is a non-native speaker of Danish. The third feature of the texts poses a more serious problem not only because Danish in contrast with English has nine progressive constructions, but also because the form most frequently used in Danish, the Simple Present, is shared by the overall structure of or viewing arrangement adopted by the text, viz. the historical present. For instance, out of the eleven progressives identified in the English source text, ten are translated into the Simple Present, type eight, and only one into a type one progressive construction in the Danish target text. See table 1 at the end of the paper which lists all nine constructions and offers ex-

[11] For the meaning of this term, see Langacker (2001).

amples of each of them. Interestingly, only constructions one, eight and nine appear in the translation. That is, construction nine only appears in two instances as translations of sentences which do not involve the progressive 'be + *-ing*' form: "They *stand* at the curb *and watch*" (DeLillo 1999a: 12) ⇒ "De *står* ved fortovskanten *og kigger*" (DeLillo 1999b: 8), "They *are* at the curbstone, *waiting*" (DeLillo 1999a: 12) ⇒ "De *står* ved kantstenen *og venter*" (DeLillo, 1999b: 8). Similar results have been obtained by Sandra Halverson (2003) who has studied a large corpus of Norwegian-English translations, namely that in the majority of cases, progressive aspect is expressed by the Simple Present tense in Norwegian. Notice that Danish and Norwegian are closely related languages. This leads to the fourth feature of the two texts, lexical aspect. If the instances of progressive aspect (read: the situational reference of a sentence) encountered in the Danish target text are solely determined by the pragmatic connections established by the reader between the situations identified by sentences, then the informational basis of such a computation must be the lexical aspect of the verb (sense 1) and/or the aspectual value compositionally derived by the syntax of the sentence in which it is found (sense 2). In other words, the frame associated with a verb, for example *run*, consider once again the description offered above, puts constraints on which connectors can be activated with the situation identified by a conventional situation description. Not only that, it also puts other constraints, consider once again the constructed example of the difference between the "Danish" frame for 'run' and the "Zambian" frame for 'run'. An example from *Underworld* (DeLillo 1999a: 11-14) is offered in the next paragraph.

The investigation of the Danish translation of DeLillo (1999a: 11-14), DeLillo (1999b: 7-10), shows that it is a very successful translation and that it has a very high quality, the main criterion being its ability to mimic the source text's potential for filmic interpretation. However, according to the tough requirements posed by the quality criterion stated in the introduction, it is of course possible to identify certain minor problems in the text. Consider once again the example mentioned at the end of section 3, paragraph (4), and its Danish translation (DeLillo 1999b) repeated/quoted below as (5a) and (5b), respectively[12]:

[12] Paragraph twelve in the source text. See the summary at the outset of the article.

(5a) Then he leaves his feet and is in the air, feeling sleek and unmussed and sort of businesslike, flying in from Kansas City with a briefcase full of bank drafts. His head is tucked, his left leg is clearing the bars. (…).

(5b) Så slipper han med fødderne og er i luften, føler sig glat og ubekymret og nærmest forretningsmæssig, kommer flyvende fra Kansas City med en mappe fuld af bankanvisnger. Han holder hovedet dukket, hans venstre ben kommer over bommene. (…).

The two possible interpretations offered by the English source text are modelled below in figures 2 and 3, respectively.

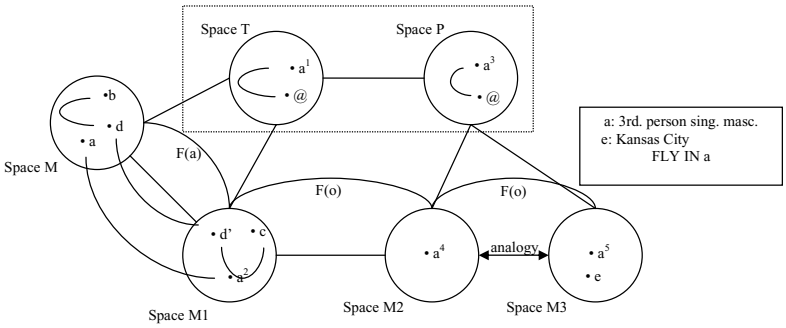

Figure 2. The protagonist's feeling interpretation of *flying in from Kansas City with a briefcase full of bank drafts*.

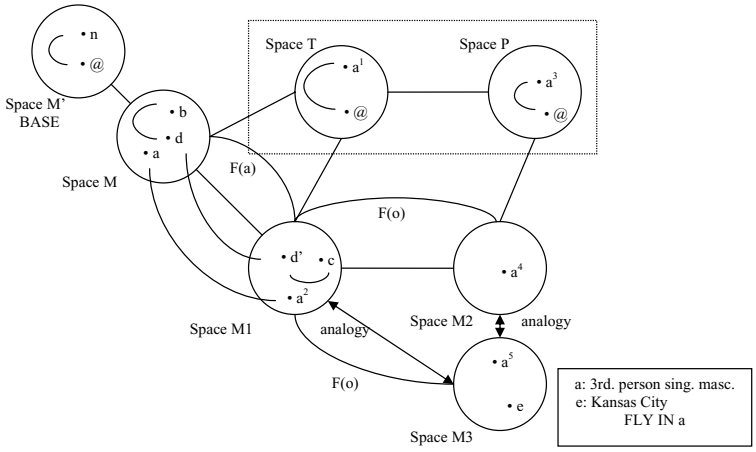

Figure 3. The non-diegetic insert interpretation of *flying in from Kansas City with a briefcase full of bank drafts*.

The first interpretation, see figure 2, is analysed as follows. The main clause of the sentence consists of two conjoined clauses. The elements activated by the first conjunct are inserted in space M (*a* = the protagonist identified by 'he'; *b* = 'his feet'; *d* = the ground). The line linking *b* and *d* in M identifies a metonymic relationship between these two entities. The elements of the second conjunct are inserted in space M1 (a' = the protagonist, c = 'in the air', d' = the ground. The line linking c and d' in M1 also identify a metonymic relationship. Idiomatically, the first conjunct also identifies a mental act which triggers the construction of a thought space, space T (the symbol '@' indicates a viewpoint role). The situations established by the two conjuncts are linked by the action-connector. The first non-finite adverbial clause of the sentence describes three perceptions which are invoked in space M2, and it triggers the construction of a perception space, space P. Since both thoughts and perceptions in literature are associated with internal speech, spaces T and P are likely to be integrated into one T/P-space, as indicated by the dotted line. The second non-finite adverbial clause of the sentence describes yet another feeling (read: T/P) of the protagonist, described in the box above M3, which is metaphoric. The double-arrow named 'analogy' shows that there exists such a metaphoric relation between

spaces M2 and M3. Since the two adverbial clauses describe properties of the subject, 'he', they both prompt object-connections.

The second interpretation, see figure 3, is analysed as follows. The MS-configuration for the main clause and the first adverbial clause of the sentence is similar to the one described in figure 2. It is only the implementation of the last adverbial clause which differs. Interpreted as a metaphorical interpretation of the agent's action, i.e. the one mentioned by the main clause, which is entertained by an omniscient narrator, the focus is shifted from spaces T and P to the base space, space M', and an object continuity is established between M1 and M3. As indicated by the double-arrow between M2 and M3, this does not imply that the establishment of an analogy relation between these two spaces is excluded.

The Danish translation, the sentence underlined in (5b), allows both interpretations. However, there is one feature which distinguishes it from its source, the verb of the first conjunct, *Så* ***slipper*** *han med fødderne og er i luften* ['Then he **releases** with his feet and is in the air']. In contrast with the English source which explicitly expresses a mental act of the protagonist, the Danish target offers an "objective" description (or medium-range shot) of his action which, albeit, due to the preceeding discourse may be interpreted as internal speech. The important question, however, is to what extent this difference between the source and target texts affects the overall interpretation. To answer this question, the next sentence in the text also needs to be taken into consideration, i.e. the sentences not underlined in (5a) and (5b). Nevertheless, consider first figure 4 which demonstrates how the role played by one's existing knowledge of film can be modelled by MS theory:

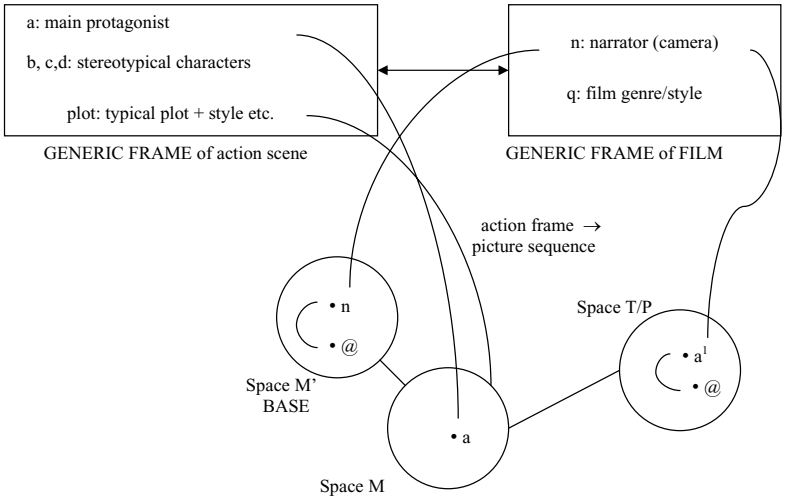

Figure 4. The recruitment of filmic frames.

The box in the upper right-hand corner of the diagram indicates that a generic frame has been activated. In contrast with a particular filmic frame which contains detailed information, for instance, of a love scene, a generic frame only contains very general information. In this case, the kind of information one would expect with any film, i.e. that it has a narrator who is associated with the camera (either an omniscient narrator or a specific character narrator who also plays a role in the story) and that it belongs to a particular style and genre. Naturally, if the object of analysis had been a particular film, more detailed filmic frames would have to be part of the analysis, i.e. the MS-configuration hypothesised. The box in the upper left-hand corner of the diagram indicates that the generic frame of action scene has been activated. Depending on which elements and relations are established in the individual spaces set up and which connections are established between these spaces, i.e. the MS configuration constructed, an action frame is selected which is consistent with the plot/style that can be aligned with it. This action frame, in turn, prompts for a particular picture sequence. The three spaces below the two boxes are identical to the three main spaces set up in the MS configuration diagrammed in figure 3, i.e. (1) the base space in which the main narrator is set up, (2) the fused T/P space in which the

protagonist narrator is set up and (3) space M in which the situation described by the source and target sentences is constructed.

The next sentence of the English source text, i.e. *His head is tucked, his left leg is clearing the bars*, is modelled in figure 5 below.

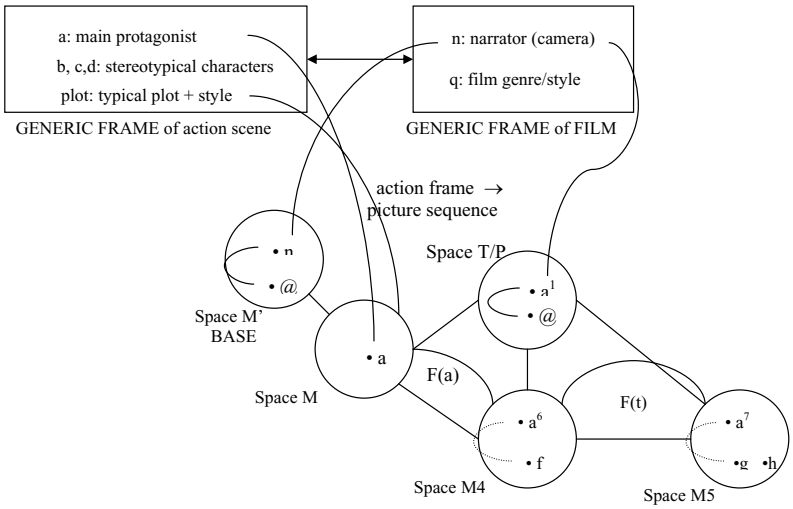

Figure 5. Constraints imposed by the Danish translation. The diagram is a compressed version of figures 3 and 4.

The elements and relations encoded by the first clause of the sentence, *His head is tucket*, are inserted in M4 (a^6 = the protagonist, *f* = 'his head', dotted line = metonymic relation). Those encoded by the second clause, *his left leg is clearing the bars*, are inserted in M5 (a^7 = the protagonist, *g* = 'his left leg', dotted line = metonymic relation, *h* = 'the bars'). Notice that the diagram both explains why the progressive form is used with the second clause, viz. due to the temporal-hierarchy connection established between M4 and M5, and why the second clause in spite of its in-progress instruction is interpreted as a full action, viz. due to the action connection established between M and M4. At the same time, the viewing arrangement established by F(a) and F(th) is compatible with a particular filmic frame, viz. that of slow-motion.

The analysis reveals that the English sentence is consistent with both interpretations. The Danish translation, on the other hand, seems

to block the second interpretation, i.e. the one compatible with a non-diegetic insert. Again, the feature responsible of the inconsistency is a verb, *Han **holder** hovedet dukket, hans venstre ben kommer over bommene* ['He **keeps** his head tucket, his left leg comes over the bars']. In contrast with the English source sentence which offers an "objective" description (or medium-range shot) of the protagonist's action, the Danish sentence seems to prompt for a "subjective" description (or close-up shot) due to the intentionality expressed by *holder* ['keeps'].

Naturally, the analysis outlined does not in itself constitute conclusive evidence that this is indeed the case. However, what the analysis demonstrates is that the MS framework provided by Fauconnier (1994, 1997) captures how knowledge structures or frames activated in memory participate in the meaning construction triggered by a text and how the text's organisation itself is reflected in the MS configuration constructed. Moreover, the continuity model proposed by Jacobsen (2004) not only explains how situation integration occurs at the lexical, sentence internal and discourse levels of conceptual organisation, but also offers a solution to how situation continuities established by system 1, 3 and 4 are mapped onto each other. The reason why this is possible is the fact that the model designed for language comprehension is assumed to be perspective-based just like Anderson's (1996) account of continuity editing in film. Finally, since the conceptual processes triggered by the text are envisioned to be implemented as a perceptual symbol systems (Barsalou 1999), it is also possible to account for the aesthetic experience generated by the text, or its "reality-like" flavour. This is so because the conceptual system devised by Barsalou (1999) relies on the same neural structures in the brain as the perceptual systems.

5. Conclusions

The analysis outlined in sections 2-4 demonstrated not only that it is possible to model all levels of organisation of a text in a consistent way, but also their interaction. This fact underlines two descriptive strengths of MS theory, (1) the flexibility and (2) and detail of analysis it offers. For instance, as regards the first, it was noted that one needs not be concerned about the specific content of the filmic frames (/ constructions) when one is modelling a linguistic text, such as the source text DeLillo (1999a: 11-14). As regards the second, the whole

point of the exercise was to establish to what extent tense and aspect play a role in the linguistic construction of an action scene, i.e. of a complex 3D simulation involving many shifts in viewpoint and tempo. For instance, the analysis of the two sentences of the source text diagrammed in figures 2-5 showed that a very high level of detail can indeed be captured by the cognitive modelling technique offered by Fauconnier (1994, 1997) and Jacobsen (2004). Therefore, the quality criterion stated in the introduction also seems viable. In fact, the type of study presented seems to offer more than that. Since the cognitive modelling technique advocated offers more precise predictions about the actual reception of a text, it might also be possible to subject the results obtained thereby to experimental testing, given that Barsalou (1999) and his colleagues manage to devise appropriate experimental tests for turning the theory of perceptual symbol systems into good science. For instance, it might be possible in the near future to construct brain studies which can establish whether the high number of 3D manipulations stipulated by the model presented in this paper has any validity. If it turns out to be the case, this would constitute evidence not only in support of the claim that system 4 plays an actual role in the text's reception but also in support of the Continuity Principle and its associated pragmatic connectors which are the basis of the theory of aspect proposed in Jacobsen (2004).

References

Anderson, Joseph D. 1996: *The Reality of Illusion. An Ecological Approach to Cognitive Film Theory*. Carbondale & Edwardsville: Southern Illinois University Press.

Barsalou, Lawrence W. 1999: Perceptual Symbol Systems. In *Behavioral and Brain Sciences*,22, 577-609.

Bordwell, David & Kristin Thompson 2001: *Film Art. An Introduction* [Sixth Edition]. New York: McGraw-Hill.

Cutrer, Michelle 1994: Time and Tense in Narratives and Everyday Language. Ph.D. Dissertation, University of California, San Diego.

DeLillo, Don 1999a: *Underworld*. London: Picador.

–, 1999b: *Underverden*. Translated into Danish by Jørgen Nielsen. Copenhagen, Denmark: Centrums Paperbacks.

Fauconnier, Gilles 1997: *Mental Spaces* (Second Edition). Cambridge: Cambridge University Press.

–, 1997: *Mappings in Thought and Language*. Cambridge: Cambridge University Press.

Glenberg, Arthur M. 1998: What Memory Is For. In *Behavioral and Brain Sciences* 20, 1-55.

Goldstone, Robert L. & Lawrence W. Barsalou 1998: Reuniting Perception and Conception. In *Cognition* 65, 231-262.

Halverson, Sandra 2003: Construal Operations in Translation. Paper presented at the *8th International Cognitive Linguistics Conference* in Logroño, Spain, 20-24 July.

Hoeg, Peter 1993: *Smilla's Sense of Snow*. Farrar, Straus and Giroux.

Høg, Peter 1993: *Frøken Smillas Fornemmelse for Sne*. Rosinante Paperbacks. Munksgaard /Rosinante, Copenhagen.

Jacobsen, Ronald R. 2004: *Situation Description (or Aspectual Underdetermination) in English. The Continuity Principle*. Ph.D. Thesis [In preparation]. Aarhus, Denmark: The Aarhus School of Business.

Langacker, Ronald W. 2001: The English Present Tense. In *English Language and Linguistics* **5** (2), 251-272.

Mourelatos, A. P. D. 1981: Events, Processes, and States. In *Syntax and Semantics, Vol. 14, Tense and Aspect*. Edited by P. Tedeschi & A. Zaenen (New York: Academic Press), pages 191-211.