

Nadine Bade & Frauke Buscher

2019

An Experimental Comparison of Two Reinterpretation
Strategies: Benefits and Challenges of Using Fictional
Contexts in Experimental Studies



In

A. Gattnar, R. Hörnig, M. Störzer & S. Featherston (Eds.)

Proceedings of Linguistic Evidence 2018: Experimental Data Drives Linguistic Theory

Tübingen: University of Tübingen

<https://publikationen.uni-tuebingen.de/xmlui/handle/10900/87132>

An Experimental Comparison of Two Reinterpretation Strategies: Benefits and Challenges of Using Fictional Contexts in Experimental Studies

Nadine Bade & Frauke Buscher

University of Tübingen

nadine.bade@uni-tuebingen.de, frauke.buscher@uni-tuebingen.de

1 Introduction

In view of the combinatorial nature of linguistic meaning, semanticists widely agree that some version of the Principle of Compositionality must certainly hold (e.g., Pagin & Westerstahl, 2011): The meaning of a complex expression is a function of the meanings of its parts and the ways they are syntactically combined. However, the conception of strict compositionality is challenged by an increasing amount of evidence that natural language interpretation is strikingly sensitive to conceptual structures and thus goes beyond a meaning construction purely based on grammar (e.g., Asher, 2011): Under certain conditions, we may (or even must) exploit information from context and world knowledge to understand a sentence. In contemporary linguistics, relevant phenomena at the semantics-pragmatics interface which require meaning adjustments by contextual knowledge are analyzed as instances of underspecification and/or coercion. The underspecification approach captures meaning adjustments by a semantic structure that is underspecified in parts and thus allows for a pragmatic specification of particular meaning components (e.g., Blutner, 1998; Dölling, 2005; Egg, 2005). Semantic underspecification occurs because either the lexical semantics of an item leaves its target argument underspecified or the compositional combination of particular items is systematically accompanied by the insertion of an underspecified variable. The potential to exploit conceptual knowledge is thus built into the semantic structure in advance and the semantic structure is considered as grammatical. The coercion approach conceives of meaning adjustments as irregular reinterpretations that are triggered by semantic conflicts arising in incompatible predicate argument combinations (e.g., Asher, 2011; Pustejovsky, 2011). This kind of combinatory conflict is solved by a linguistic adaptation mechanism that paves the way for exploiting on conceptual knowledge to save the interpretation of the conflicting terms. The potential to adjust the meaning is thus added in individual cases that require for it. In this paper, we focus on meaning adjustments of the second type illustrated by well-known phenomena like Complement Coercion, Aspectual Coercion, Metonymy and Personification, cf. (1).

- | | |
|---|---------------------|
| (1) a. <i>The author began the book.</i> | Complement Coercion |
| b. <i>Peter sneezed for five minutes.</i> | Aspectual Coercion |
| c. <i>The cappuccino wants to pay.</i> | Metonymy |
| d. <i>The storm raged all night long.</i> | Personification |

For example, in (1a), a combinatory conflict arises since the argument *the book* denoting a physical object does not meet the requirements of the predicate *begin* selecting for an event as internal argument. This semantic mismatch triggers an adaptation mechanism that allows us to infer an event like writing or reading the book. Thus, the adaptation mechanism makes the

sentence interpretable despite the combinatory conflict. In a similar way, the semantic mismatches between *sneeze* and *for five minutes*, *the cappuccino* and *want*, *the storm* and *rage* trigger reinterpretations.

Besides the loose definition as linguistic repair mechanism, current conceptions of coercion differ significantly. There is not a unique way to deal with coercion in theoretical linguistics, but several types of adaptation mechanisms are proposed (for some theoretical approaches to these phenomena, see, e.g., Asher (2011) and Pustejovsky (2011); for an overview, see de Swart (2011)). For example, in (1a), coercion can be modeled as interpolation of an underspecified event variable that allows for inferring an adequate event. In (b), coercion can be defined as an adjustment of an event which is introduced by the verb, resulting in an interpretation of the sneezing as an iterative event. In (c), coercion can be explained by a referential adjustment that shifts the reference of *the cappuccino* from the beverage to the person who drank the cappuccino. In (d), coercion can be described as a conceptual adjustment that attributes human traits to an inanimate entity.

Psycholinguistic research has provided broad empirical evidence for the assumption that meaning adaptations are cognitively real. The interpretation of the examples in (1) which are distinguished theoretically leads to increased processing costs compared to a straightforward compositional interpretation as shown by a variety of experimental studies using a broad spectrum of methods, including reading times (e.g., see Baggio et al., 2009; Frisson & McElree; Kuperberg et al., 2010; McElree et al., 2001; McElree et al., 2006; Pyllkkänen et al., 2009; Pyllkkänen & McElree, 2007; Scheepers et al., 2008; Traxler et al., 2002; Traxler et al., 2005; 2008, for complement coercion; Bott, 2010; Brennan & Pyllkkänen, 2008; Piñango et al., 1999; Piñango et al., 2006; Todorova et al., 2000, for aspectual coercion; Schumacher & Weiland, 2011; Schumacher, 2011, 2013, 2014; Weiland et al., 2014, for metonymy; Filik, 2008; Filik & Leuthold, 2008; Nieuwland & van Berkum, 2006, for personification). The assumption of adaptation mechanisms at the semantics-pragmatics interface is not a mere matter of abstract theoretical reasoning but can be proved empirically.

However, studies have mostly tested individual phenomena and thus, isolated adaptation mechanisms. Up to now, there is no systematic comparison of different types of adaptation mechanisms and it is still an open question whether their theoretical distinction is cognitively real or not. This paper addresses the empirical foundation of the theoretical distinction of several types of adaptation mechanisms: Are they just due to different theoretical backgrounds, or can they be verified on empirical grounds? An answer to this question promises important implications for theoretical analyses of coercion.

As a starting point, our studies focused on comparing interpolation, cf. (1a), and personification, cf. (1d). In order to see whether and in which contexts evidence for both adaptation mechanisms can be found, we conducted a self-paced reading study exploring German mental attitude adverbials like *absichtlich* ‘intentionally’. In our study, we combined the methods and perspectives of psycholinguistics and literary scholarship to settle the question outlined. Therefore, this paper also has a methodological concern. It shows how using fictional contexts allows for a systematic experimental comparison of two adaptation mechanisms in one and the same linguistic expression and discusses how fictional contexts are a useful resource for investigating coercion empirically.

The paper is structured as follows. Section 2 outlines the issue we address by presenting an expression which allows for different adjustment operations. We introduce German mental attitude adverbials as the case in point and illustrate what the properties of interpolation and personification are in more detail. Moreover, Section 2 argues that fictional contexts are a useful tool for creating sophisticated experimental designs which allow us to investigate empirically the full range of adaptation mechanisms. Section 3 presents a self-paced reading study including

two pretests (plausibility rating study and acceptability rating study) and discusses the methodological challenges of using fictional contexts in experimental set-ups. Section 4 concludes and outlines some open questions.

2 The issue

One major challenge with comparing interpolation and personification experimentally is to find a phenomenon that allows solving a combinatory conflict with the help of both adaptation mechanisms. German mental attitude adverbials – or more precisely, their subtype intentional adverbials (cf. Buscher, 2013, 2018)¹ – turn out to be an adequate case study.

- (2) *Der Mann steht absichtlich im Schatten.*
 the man stands intentionally in.the shade
 ‘The man is standing in the shade intentionally.’

Intentional adverbials such as *absichtlich* ‘intentionally’ describe that the event introduced by the verb was initiated and that the initiator of the event has an intention (Buscher, 2018). In (2), the man is interpreted as the attitude holder. (2) can be paraphrased as ‘There is an event *e* where the man *m* stands in the shade. At the same time, there is a state *z* in which the man *m* who initiated the event *e* has the intention *r* that he has the property of standing in the shade.’ Accordingly, intentional adverbials require an anchor argument which refers to an animate referent who is both able to initiate an event and to have an attitude. If the anchor argument is filled by an adequate attitude holder, such as the man in (2), composition proceeds smoothly (see Buscher, 2018, for details). In contrast to (2), the example in (3) does not offer an appropriate anchor argument:

- (3) [non-fictional context, e.g., everyday conversation or newspaper]
Das Auto steht absichtlich im Schatten.
 the car stands intentionally in.the shade
 ‘The car is standing in the shade intentionally.’

The anchor argument *das Auto* ‘the car’, which is introduced linguistically, refers to an inanimate entity and thus, does not meet the selectional restrictions of *absichtlich* ‘intentionally’. A straightforward compositional interpretation of (3) results in a semantic mismatch between *absichtlich* ‘intentionally’ and *das Auto* ‘the car’. However, intentional adverbials allow a pragmatic identification of the attitude holder (see Buscher, 2013, 2018; Eckardt, 2003: 264; Pittner, 2004: 284, fn. 12 for German and Dowty, 2007: 62, fn. 29 for English). In (3), *absichtlich* ‘intentionally’ describes the intention of a person inferred from the context who parked the car. (3) can be paraphrased as ‘There is an event *e* where the car *c* stands in the shade. At the same

¹ The possibility of pragmatic identification is quite restricted and not available for all mental attitude adverbials:

- (i) **Das Auto steht freiwillig im Schatten.*
 the car stands voluntarily in.the shade
 ‘The car is standing in the shade voluntarily.’

The different options available to identify the attitude holder are related to an independently motivated subclassification of mental attitude adverbials (Buscher 2013, 2018): intentional adverbials such as *absichtlich* ‘intentionally’ permit a pragmatic identification of the attitude holder, while assimilative adverbials such as *freiwillig* ‘voluntarily’ strictly adhere to the principle of compositionality and do not permit a pragmatic identification of the attitude holder. Based on this observation, Buscher (2018) developed a semantic analysis of intentional adverbials that derives the compositional interpretation as well as the coercion case on the basis of a single lexical entry for *absichtlich* ‘intentionally’. The analysis is spelled out formally within Asher’s (2011) framework of a type-driven context-sensitive lexical semantics.

time, there is a state z in which a person i who initiated the event e has the intention r that the car c has the property of standing in the shade.’

Based on linguistic analysis and empirical studies, Buscher (2018) shows that the pragmatic identification of the attitude holder is neither based on lexically anchored underspecification of intentional adverbials nor on different syntactical positions of intentional adverbials. The pragmatic identification of the attitude holder follows from a meaning adaptation that is equivalent to the meaning adaptation in Complement Coercion (see Buscher (2018) for an in-depth theoretical comparison of the meaning adaptations occurring with *begin* and *intentionally*): The semantic mismatch between *absichtlich* ‘intentionally’ and *das Auto* ‘the car’ is solved by interpolating an additional underspecified variable (cf. i above) to the semantic representation of (3). The underspecified variable represents the initiator of the event. It allows inferring the initiator and thus, the attitude holder. Most importantly, in everyday speech, the interpolation outlined is necessary to interpret the example in (3). There is no other option available to interpret the sentence.

Interpolation is characterized by three main aspects (Buscher (2018); see Asher (2011) for the same results regarding Complement Coercion): (i) It is a non-destructive operation and leaves the denotation of the anchor argument unchanged: *Das Auto* ‘the car’ still refers to an inanimate entity and it is not interpreted as an entity that is able to have an attitude. (ii) It solves the semantic mismatch locally: It is still the car that is localized in the shade, not the attitude holder inferred. (iii) It is an adaptation mechanism which is licensed linguistically: The interpolation is anchored in the lexical entry of *absichtlich* ‘intentionally’. Thus, the availability of interpolation depends on the lexical material involved in a semantic mismatch – but it does not depend on specific text types, i.e., it is generally available in everyday speech (provided that it is plausible that the described situation was initiated and an initiator can be inferred).

However, literary scholarship suggests that context and text type can have an important impact on the availability of certain (!) adaptation mechanisms. They show that fictional texts offer additional possibilities to solve semantic mismatches since, unlike utterances in everyday speech, they do not claim to make statements about the actual world and thus are not limited by conceptual restrictions thereof. This is illustrated by the literary example in (4), taken from the poem ‘My life had stood a loaded gun’ by Emily Dickinson (J754; Dickinson, 1955: 574).

(4) The mountains straight reply.

What (4) shares with the other examples outlined above is that a compositional interpretation leads to a semantic mismatch. The verb *reply* comes with a selectional restriction limiting possible arguments for *reply* to animate and speaking individuals. The argument *the mountains* refers to inanimate entities and thus, does not meet the requirements of the verb. As shown by literary scholarship, it is still possible to reinterpret (4) in two ways (Bauer et al., 2015): First, the sentence could be read metaphorically in that the mountains are producing an echo. In this case, the meaning of the verb *reply* and its selectional restrictions are shifted. It is understood as an acoustic reaction of any kind. As a result, the selectional restrictions of *reply* become more flexible and arguments are not restricted to human beings. In addition, there is a second, less straightforward way to reinterpret (4) which hinges on the fictional nature of the text the example is embedded in. This adaptation mechanism shifts the meaning of *the mountains* and changes our conceptual knowledge regarding the referent thereof, i.e., the property of being animate is ascribed to the mountains so that they can literally answer.

At first glance, the data suggest that the availability of the second interpretation, called personification in the following discussion, is dependent on the fictional nature of the context. The possibility of ascribing the property of being animate to an inanimate entity is mainly available in fiction and implausible in everyday conversation due to conceptual restrictions. However, personifications can also be used in non-fictional contexts as everyday conversations, cf.

(1d), and often, they can be found in speeches, headlines and commercials.² This observation provides evidence that this adaptation mechanism is part of the grammatical system; the grammar used in poetry must be considered as a variety of a non-poetic language, and not as a proper language in itself (Bade & Beck, 2017). Thus, evidence from fictional texts is valid and valuable for linguistic theory. Nevertheless, the limits of interpretational flexibility are broader in fictional contexts than in non-fictional contexts. Examples occurring in poetry may display the full range of reinterpretations the grammar allows for since they are not restricted by conceptual knowledge.

Against this background, we predict a sentence like (3) to become ambiguous in a fictional context, e.g., in a poem, a fairy tale or a cartoon movie like ‘Cars’ (Pixar), cf. (5):

- (5) [fictional context, e.g., poem, fairy tale or cartoon movie]
Das Auto steht absichtlich im Schatten.
 the car stands intentionally in.the shade
 ‘The car is standing in the shade intentionally.’

The semantic mismatch between *absichtlich* ‘intentionally’ and *das Auto* ‘the car’ can be solved either by interpolation (as licensed by the lexical entry of *absichtlich* ‘intentionally’) or by personification (as licensed by the fictional nature of the context). If an appropriate context for a personification exists, the anchor argument *das Auto* ‘the car’ can be taken into account as the attitude holder. The car can be reinterpreted as animate and thus, as the bearer of the intention. Therefore, fictional contexts prove to be a valuable methodological resource for comparing different types of adaptation mechanisms experimentally. They allow us to manipulate the availability of certain adaptation mechanisms. Using fictional contexts, we are able to compare both types of meaning adaptations within one and the same linguistic expression.

Compared to interpolation, personification is characterized by the following aspects from a theoretical point of view (see Table 1 for an overview): (i) In contrast to interpolation, personification does not enrich the semantic representation. It changes the denotation of the anchor argument: *Das Auto* ‘the car’ does not refer to an inanimate entity, but is interpreted as an entity that is able to have an attitude. Therefore, personification is destructive in nature. (ii) As interpolation, personification solves the semantic mismatch locally: In (5), it is still the car that is localized in the shade. (iii) In contrast to interpolation, personification is not licensed by the lexical material involved but by certain characteristics of the text.

Table 1. Characteristics of interpolation and personification

	interpolation	personification
semantic enrichment	yes	no
destructive	no	yes
local	yes	yes
lexically anchored	yes	no

Provided that our theoretical assumptions are correct and interpolation and personification are very different in nature, these adaptation mechanisms should involve different processing costs if their distinction is also cognitively real. Specifically, we expect the destructive adaptation

² However, usually personifications appear in a less radical form in these types of texts. The inanimate entities are not – as often the case in fictional texts – interpreted as animate in the strict sense, cf. the storm in (1d). Rather certain properties of animate subjects are transferred to the inanimate entity. Our conceptual knowledge is kept in touch for the most part, i.e., the limits for adjustment of conceptual knowledge are less flexible.

mechanism, i.e., personification, to be more costly because interpolation is linguistically encoded whereas personification is contextually licensed.

Against this background, the following questions have to be discussed: (i) Do experimental data provide evidence for the assumed influence of the context on the choice of adaptation mechanisms? More precisely: Is interpolation preferred in a non-fictional context? And is it possible to make personification more prominent with providing fictional contexts? (ii) Is the theoretical distinction between interpolation and personification cognitively real?

To answer these questions, we conducted a self-paced reading study and two pretests.

3 Self-paced reading study

3.1 Core idea

The self-paced reading study we conducted aimed at testing if interpolation and personification cause increased processing costs compared to a straightforward compositional interpretation. In addition, we wanted to see if these extra processing costs differ significantly between the two mechanisms, thereby drawing on our initial questions whether the two mechanisms can be distinguished on an empirical basis. We tested sentences in which the subject (animate referent vs. inanimate referent) matches or mismatches the intentional adverbial, and enforced two different types of adaptation mechanisms by manipulating the context (non-fictional vs. fictional), cf. (6) and (7). The slashes indicate regions into which the target sentence was split for representation.

- (6) [non-fictional vs. fictional context]
Das Auto / steht / absichtlich / im Schatten.
 the car stands intentionally in.the shade
 ‘The car is standing in the shade intentionally.’
- (7) [non-fictional vs. fictional context]
Der Mann / steht / absichtlich / im Schatten.
 the man stands intentionally in.the shade
 ‘The man is standing in the shade intentionally.’

Target sentences with animate subjects allow for a straightforward compositional interpretation, irrespective of context. They served as control conditions. Target sentences with inanimate subjects require, irrespective of context, a meaning adaptation according to our assumptions. As illustrated above, we expected that a target sentence with an inanimate subject presented in a non-fictional context can only be interpreted by inferring an attitude holder. In a fictional context it is possible to assume that the car itself is animate and holds the attitude. To be able to compare the cognitive load associated with each adaptation mechanism we measured the reading times on the region disclosing the semantic mismatch (Region 3), i.e., the adverbial, and the following PP (Region 4). Extra processing costs for the critical conditions compared to control conditions should be reflected in an increase in reading times on the adverbial and/or the PP.

There is a crucial methodological challenge that comes with this set-up: Even when we observe different reading time patterns for (6) depending on the context, we cannot be sure which type of adaptation mechanism was chosen and thus, is responsible for the reading time pattern on the adverbial and/or the PP. To be able to see what adaptation mechanism was cho-

sen, we drew on findings of previous psycholinguistic studies taking into consideration fictionality of the context as a factor in investigating personification.³ These studies found that processing difficulties that come along with personification disappear if the context is strong enough. For example, Filik (2008) and Filik & Leuthold (2008) tested personifications as in (8) using eye-tracking and ERP.

(8) The cat picked up the chainsaw.

They find that the N400 effect usually found for reinterpretation disappears on the object if a ‘Tom & Jerry’ scenario is established as the context. This suggests that the adjustment of world knowledge – i.e., assuming that cats can pick up things like chainsaws – happens immediately. In a similar spirit, a famous study by Nieuwland & van Berkum (2006) finds that the sentence in (9) does not yield increased processing costs in contexts where the peanut is established as animate before.

(9) The peanut was in love.

The authors conclude that lexical/semantic violations can be overruled by context immediately. As opposed to the studies by Filik and colleagues the work by Nieuwland & van Berkum (2006) furthermore suggests that a context can be established as fictional by using personifications, i.e., participants need not be familiar with the referents in order to accept world knowledge adjustments.

Based on these studies, we expected extra processing costs that come along with the reinterpretation of the car as animate to disappear if the car is already introduced as animate. Therefore, we included a second conjunct as a control measure for which adaptation mechanism was chosen in the first conjunct (cf. *und beobachtet das Wettrennen* ‘and watching the race’).

(10) [non-fictional vs. fictional context]

Das Auto /steht /absichtlich /im Schatten /und /beobachtet das Rennen.
the car stands intentionally in.the shade and watches the race
‘The car is standing in the shade intentionally and watching the race.’

(11) [non-fictional vs. fictional context]

Der Mann /steht /absichtlich /im Schatten /und /beobachtet das Rennen.
the man stands intentionally in.the shade and watches the race
‘The man is standing in the shade intentionally and watching the race.’

The second conjunct incorporating an intentional verb ultimately enforced a reading where the car has to be taken to be animate, i.e., a personification. Depending on what kind of adaptation mechanism was triggered and used in the first conjunct, the reading times should increase or not on the second conjunct. If the car was already reinterpreted as animate in the first conjunct, *das Auto* ‘the car’ now is compatible with *beobachten* ‘to watch’ and the reading times on the second conjunct should not increase. In contrast, if an attitude holder was inferred in the first conjunct, *das Auto* ‘the car’ is still interpreted as inanimate and thus, it is incompatible with *beobachten* ‘to watch’. In this case, the participant should recognize in the second conjunct that the reinterpretation strategy has to be changed and that the car has to be reinterpreted as animate. Thus, in the non-fictional context, the reading times on the second conjunct should increase significantly compared to the reading times in the fictional context since the initial interpretation has to be revised. In the control conditions, no processing difficulties should arise since a compositional interpretation of the second conjunct is possible in both contexts. The reading

³ As mentioned before, a systematic comparison of different types of coercion processes and how they are affected by context is missing in the literature we are aware of.

time pattern we expected is schematized in Figure 1 below; the straight line symbolizes ‘no extra processing costs’ and the peaks symbolize ‘extra processing costs’.

To sum up, our design has the advantage that the predictions regarding interpolation and personification correspond to different reading time patterns. If the expectations outlined turn out to be true, we already have evidence that the two adaptation mechanisms are cognitively real. Moreover, by using the set-up described we can infer from the reading time pattern found in the second conjunct what kind of adaptation mechanism was chosen in the first conjunct. Therefore, we are in the position to see whether their processing costs are different by comparing the reading times on the adverbial and/or the PP in the first conjunct.

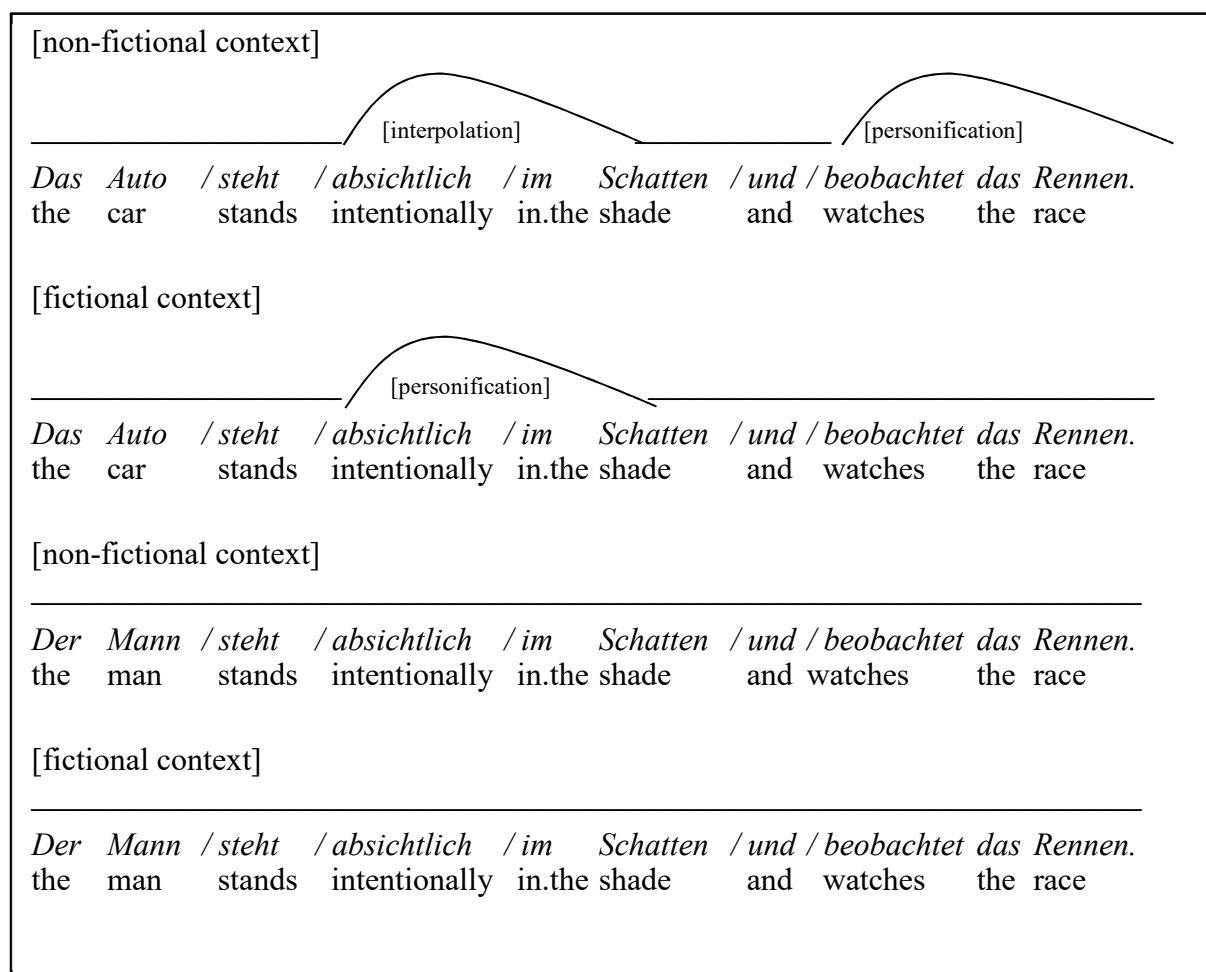


Figure 1. Schematized expectations for reading time patterns

3.2 Materials

We employed a 2×2 design, with both factors SUBJECT (animate referent vs. inanimate referent) and CONTEXT (non-fictional vs. fictional) being manipulated within participants and within items. We crossed the two-level factor SUBJECT with the two-level factor CONTEXT; materials consisted of 36 experimental items that were realized in all four conditions. (12) and (13) show the four variants of a sample item taken from our materials. The bold expressions point out the parts we manipulated in the contexts and target sentences. The slashes indicate regions into which the target sentence was split for representation. The translated versions of the contexts can be found in (15) below.

(12) [non-fictional context]

*Eine Tante liest ihrer Nichte **eine Nachricht aus der Albstädter Zeitung** vor: Auf einer kleinen Berghütte in den Alpen herrschte am Weihnachtsmorgen eine friedliche Stimmung. Die Magd fütterte die Kühe, während **der Knecht** die Pferde striegelte. Die Bäuerin bereitete den Weihnachtsbraten zu und **der Bauer** heizte den Kachelofen ein. Die gute Stube war zu diesem Zeitpunkt noch nicht geschmückt.*

[target with inanimate subject]

Die Tanne / lag / *absichtlich* / im Heuschober / und / *schwänzte* die Arbeit.
the tree lay intentionally in.the haystack and avoided the work
'The tree was lying in the haystack intentionally and avoiding work.'

[target with animate subject]

Der Bauerssohn / lag / *absichtlich* / im Heuschober / und / *schwänzte* die Arbeit.
the farmer's son lay intentionally in.the haystack and avoided the work
'The farmer's son was lying in the haystack intentionally and avoiding work.'

(13) [fictional context]

*Eine Tante liest ihrer Nichte **eine Geschichte aus einem Märchenbuch** vor: Auf einer kleinen Berghütte in den Alpen herrschte am Weihnachtsmorgen eine friedliche Stimmung. Die Magd fütterte die Kühe, während **der Kamm** die Pferde striegelte. Die Bäuerin bereitete den Weihnachtsbraten zu und **der Blasebalg** heizte den Kachelofen ein. Die gute Stube war zu diesem Zeitpunkt noch nicht geschmückt.*

[target with inanimate subject]

Die Tanne / lag / *absichtlich* / im Heuschober / und / *schwänzte* die Arbeit.
the tree lay intentionally in.the haystack and avoided the work
'The tree was lying in the haystack intentionally and avoiding work.'

[target with animate subject]

Der Bauerssohn / lag / *absichtlich* / im Heuschober / und / *schwänzte* die Arbeit.
the farmer's son lay intentionally in.the haystack and avoided the work
'The farmer's son was lying in the haystack intentionally and avoiding work.'

The target sentences consisted of a definite NP functioning as subject (Region 1), followed by a verb in the past tense (Region 2), then an intentional adverbial (Region 3), a PP functioning as locative modifier (Region 4), the conjunction *und* 'and' (Region 5) and finally a second conjunct introducing a VP with an intentional transitive verb and a definite object (Region 6). The regions of interest were the Region 3 indicating the match / mismatch between the subject and the adverbial, the spill-over Region 4 and the Region 6 indicating that personification is the required adaptation mechanism.

- (14) a. *stehen* 'to stand', *liegen* 'to lie', *lehnen* 'to lean', *hängen* 'to hang', *klemmen* 'to clamp', *stecken* 'to be stuck'
b. *absichtlich* 'intentionally', *vorsorglich* 'precautionarily', *bewusst* 'consciously', *irrtümlich* 'mistakenly', *unabsichtlich* 'unintentionally', *versehentlich* 'inadvertently'

As subject we used one of the 36 NPs identified as the most plausible for a fictional context by the plausibility rating we ran as a pretest (see 3.3.1). As verb we always used one of the positional state verbs listed in (14a). Each verb was used in six target sentences. As intentional adverbial we used one of the adverbials listed in (14b). Three of them describe a positive attitude (*absichtlich* 'intentionally', *vorsorglich* 'precautionarily', *bewusst* 'consciously'), three of them describe a negative attitude (*irrtümlich* 'mistakenly', *unabsichtlich* 'unintentionally', *versehentlich* 'inadvertently'). Each adverbial was used in six target sentences and each verb

was combined with each adverbial, i.e., no verb-adverbial combination appeared more than once. We want to illustrate the features of the context using the English translations, cf. (15).

(15) a. [non-fictional context]

*An aunt is reading her niece a story from the local newspaper: It was Christmas morning in a small alpine hut and the mood was peaceful. A milk maid was feeding the cows, while **the stable boy** was brushing the horses. The farmer's wife was cooking christmas dinner while **the farmer** was lighting the wood stove. The Christmas decorations had not been hung yet.*

b. [fictional context]

*An aunt is reading her niece a fairytale: It was Christmas morning in a small alpine hut and the mood was peaceful. A milk maid was feeding the cows, while **the comb** was brushing the horses. The farmer's wife was cooking christmas dinner while **the bag bellow** was lighting the wood stove. The Christmas decorations had not been hung yet.*

We used rather long contexts which were completely parallel in the fictional and non-fictional case except for three points where they diverged. We had five sentences in the context in total, with the first one being an introductory sentence establishing whether we are in a newspaper or fairytale setting. The second sentence introduced the overall situation, but did not introduce any referents yet. In the third and fourth sentence two participants were introduced, the second of which was inanimate or animate depending on the context condition. Then there was a rather neutral sentence again which made the connection to the following target sentence.

The experimental items were combined with 48 filler items, whose general shape was similar to our experimental items, i.e., we used contexts with five sentences and one critical target. We used four types of filler items serving different purposes; each type included 12 filler items (six filler items with fictional context and six filler items with non-fictional context; each type is illustrated by an example with fictional context). The well-formedness of the filler items varied from completely acceptable to completely unacceptable; ill-formed filler items contained grammatical violations or conceptual violations.⁴ The first type of fillers we used served the purpose of obscuring the general aim of the study. The target sentence within this type of filler neither contained a grammatical violation nor conceptual conflict. It was straightforwardly interpretable and acceptable. Fictionality was invoked, however not by using personification.⁵ The target sentence contained an adverbial different from intentional adverbials.

(16) *Eine Tante liest ihrem Neffen aus einem Comic vor: Auf dem Empire-State-Building fand ein Treffen von Superhelden statt. Wonderwoman berichtete von einem neuen Bösewicht und Batman erzählte von dessen geheimer Waffe. Spiderman lieferte Informationen über einen geplanten Anschlag, während Captain America einen Rettungsplan entwickelte. In diesem Moment zog ein massives Gewitter auf.*

Die Wolkenfront /bedeckte /schlagartig /die Sonne /und /verdunkelte den Himmel.

‘An aunt is reading her nephew a story from a comic book. There was a meeting of super heroes on top of the Empire State Building once. Wonderwoman was telling the others about a villain and Batman reported about his secret weapon. Spiderman delivered information regarding a planned attack while Captain America developed a counter attack. At this moment a storm started. The clouds suddenly covered the sun and darkened the sky.’

⁴ We took as a background results of an unpublished study testing these grammatical and conceptual violations (conducted by Carolin Dudschig, Barbara Kaup & Claudia Maienborn, p.c.).

⁵ Due to the nature of our material translations must be considered as approximations to the original German material.

The second type of filler items ensured that interpolation was a successful strategy throughout the whole target sentence some of the time. Since our experimental items with inanimate subjects were such that interpolation was possible in the first conjunct but never in the second we were suspecting that participants might drop this strategy at some point since it had to always lead to revision. To counterbalance this effect, we thus included 12 filler items containing intentional adverbials where interpolation was possible in the first conjunct and the resulting reading was also compatible with the second conjunct.

Analogously to our experimental items there was a conflict between the intentional adverbial and the subject in the first part of the target sentence. However, since the context did not contain any personifications (i.e., inanimate entities being depicted as animate) interpolation was considered as the more likely strategy. In contrast to the experimental items, the second conjunct did not require an adjustment of conceptual knowledge, i.e., interpolation could be used as a strategy consistently within the sentence. These fillers, as well, were assumed to be completely acceptable.

- (17) *Eine Großmutter liest ihrem Enkel aus einem Fantasybuch vor: In Mitteleerde herrschte wieder ein erbitterter Krieg um den sagenumwobenen Ring. Der Elb schoss mit Pfeil und Bogen, während der Zwerg mit einer Axt um sich schlug. Der Zauberer verteidigte das Dorf und der Hobbit ritt hilfesuchend zum Schloss. Doch all die Kämpfe waren vergebens.*

Der Ring /steckte /versehentlich /in einer Felsspalte /und /war verloren.

‘A grandmother is reading her grandson a story from a fantasy book. A bitter dispute about the famous ring started in Middle Earth. The Elf was shooting with bow and arrow, while the dwarf used his axe. The magician defended the village, and the hobbit rode to the castle to seek help. But all fighting was in vain: The ring got involuntarily stuck in a crevice and was lost.’

The third type of filler items was meant to counterbalance the effect of the semantic conflict evoked in the experimental items (i.e., the semantic conflict between the intentional adverbial and the subject). The target sentences of these filler items involved a semantic conflict between a verb selecting for a plural noun phrase (e.g., *verstreuen* ‘disperse’) and a noun phrase in the singular (e.g., *den Kieselstein* ‘the pebble’). In contrast to the experimental items this kind of semantic conflict is not resolvable through adaptation mechanisms. As a result, these filler items were always unacceptable.

- (18) *Ein Vater liest seiner Tochter aus einem Kinderbuch vor: In einem verwunschenen Schlossgarten geschahen zauberhafte Dinge. Die Zauberin verhexte die Blumenerde und der Gärtner verwandelte die Rosen in Tulpen. Die Königin badete im Jungbrunnen, während die Prinzessin mit den Fröschen spielte. Doch einer fand in diesem idyllischen Garten keinen Frieden.*

Der Diener /verstreute /missmutig /den Kieselstein /und /verschmähte die Gartenarbeit.

‘A father is reading his daughter a story from a children’s book: In an enchanted garden of the castle magical things happened. The wizard bewitched the soil and the gardener turned the roses into tulips. The queen was bathing in a fountain of youth and the princess was playing with frogs. All but one found peace in the magical garden. The servant sullenly dispersed the pebble and tried to avoid the work in the garden.’

The fourth type of filler items was meant to counterbalance the effect of the conceptual conflict evoked in the experimental items (i.e., inanimate entities as animate entities). In these filler items, grammatical restrictions are observed (plural noun with verb selecting plural), however there was a conceptual conflict between the verb and the direct object. In contrast to our experimental items, the conflict could not be resolved through meaning adaptation. Since the context

did not license personification (i.e., no inanimate subjects were depicted as animate), personification was not an option for resolution in the target either, and made it unacceptable.

- (19) *Ein Vater liest seiner Tochter eine Legende vor: In einem kleinen Märchenwald hinter den Bergen geschah etwas Furchtbares. Der böse Zauberer belegte den Wald mit einem Regenfluch und die hämische Nixe verhexte den Fluss. Der Troll ertrank in seiner Höhle, während sich die Fee auf einen Baum retten konnte. Es war ein trauriger Tag. Die Elfe /trocknete /deprimiert /die Pfützen /und /verteufelte den Regen.*

‘A father is reading his daughter a legend: In a magical forest in the mountains something horrible had happened. The evil wizard put a rain curse on the forest and the gleeful mermaid bewitched the river. The troll drowned in a cave, while the fairy could escape on a tree. It was a sad day. The elf was drying the puddles and wished the rain to hell.’

The experimental items were distributed over four lists. The four lists were counterbalanced across items and conditions: Each list included only one version of each experimental item and each list contained nine experimental items in each of the four conditions. Experimental items were intermixed with 48 filler items. We pseudo-randomized the sequence of the items so that no two experimental items followed each other, and no two fillers of the same sort followed each other (i.e., two out of each type of filler specified above). Based on this guideline, experimental items and filler items were randomized separately for each participant by E-Prime.

3.3 Pretests on plausibility and acceptability

There were several uncertainties regarding the material due to the few previous studies we could rely on. First, we suspected that only a certain kind of inanimate entities can plausibly be interpreted as animate, independent of context. This is why we did a first pretest on plausibility of inanimate entities as being animate to identify the most plausible referents for fictional contexts. Second, we wanted to make sure that the context manipulation had the desired effect and that our contexts were overall acceptable. Thus, we did a second pretest on the acceptability of our target sentences in the given contexts.

3.3.1 Pretest I: Plausibility rating of (in)animate subjects

In order to control for how plausible it is that a given inanimate entity is interpreted as animate we ran a simple pen-and-paper pretest where participants were given 103 nouns referring to inanimate entities in one of three different randomizations and were asked to answer the question ‘How plausible do you think it is that this item is presented as animate in a fictional context?’. They were asked to judge the plausibility of the noun on a scale from 7 (‘very plausible’) to 1 (‘not plausible’). 24 students from the University of Tübingen participated in the study. They were all native speakers of German.

We took the mean plausibility scores for the analysis (the z-transformed scores per participant) and selected the 36 most plausible items based on these scores. The chosen items ordered by decreasing plausibility are given in (20). The best rated item was *Puppe* ‘doll’ and had an average rating of 6.7, the least plausible item was *Cello* ‘cello’ and had an average rating of 4.21. Thus, all the items we ended up using, were plausible above average.

- (20) *Puppe* ‘doll’, *Pfefferkuchenmann* ‘gingerbread man’, *Vogelscheuche* ‘scarecrow’, *Teddybär* ‘teddy’, *Schneemann* ‘snowman’, *Zinnsoldat* ‘tin soldier’, *Gartenzwerg* ‘garden gnome’, *Spiegel* ‘mirror’, *Besen* ‘broom’, *Lokomotive* ‘loco’, *Auto* ‘car’, *Hexenbuch* ‘magic book’, *Teekanne* ‘teapot’, *Eiche* ‘oak’, *Trauerweide* ‘weeping willow’, *Schaukelpferd* ‘rocking horse’, *Spieluhr* ‘musical clock’, *Tanne* ‘christmas tree’, *Bus* ‘bus’, *Stiefel* ‘boots’, *Koffer* ‘suitcase’, *Sonnenblume* ‘sunflower’, *Standuhr* ‘hall clock’,

Tasse ‘cup’, *Teppich* ‘carpet’, *Bulldozer* ‘bulldozer’, *Sparschwein* ‘piggy bank’, *Geige* ‘violin’, *Schwert* ‘sword’, *Kerzenständer* ‘candlestick’, *Füllfederhalter* ‘pen’, *Stehlampe* ‘floor lamp’, *Apfelbaum* ‘apple tree’, *Klavier* ‘piano’, *Hammer* ‘hammer’, *Cello* ‘cello’

Each of these 36 nouns were used as subject in one of the items for the self-paced reading study.

3.3.2 Pretest II: Acceptability rating study

To see whether our fictional contexts worked in general (meaning they make the relevant personification more prominent) we used the material created for the self-paced reading study for a second pretest: an acceptability rating study.

Materials

The material for the acceptability rating study was identical to the material for the self-paced reading study outlined above (see 3.2). The experimental items were distributed over four lists. The four lists were counterbalanced across items and conditions: Each list included only one version of each experimental item and each list contained nine experimental items in each of the four conditions. Experimental items were intermixed with 36 filler items. We used the same filler types described above, however without the type that made interpolation possible throughout.

Participants

24 students from the University of Tübingen participated in the study and received a monetary reimbursement (12 Euros). They were all native speakers of German and were naïve regarding the purpose of the study.

Procedure

The acceptability rating study also was a pen-and-paper study. Participants were asked to judge the acceptability of the target sentence in the given context on a scale from 5 (‘highly acceptable’) to 1 (‘not acceptable’). Moreover, they were asked to answer comprehension questions after every five to seven items to make sure they were involved in the task. Participants did the experiment in three blocks with ten minute breaks in between.

Predictions

Based on the theoretical assumptions outlined above, we expected an interaction between SUBJECT and CONTEXT. In non-fictional contexts, target sentences with an inanimate subject should only allow for the interpolation of an attitude holder.

If the non-fictional context only allows for interpolation, but does not license personification, the acceptability of the target sentences with inanimate subjects in non-fictional contexts should be significantly lower than target sentences with animate subjects in non-fictional contexts. If the fictional context licenses personification the acceptability of target sentences with inanimate subjects in fictional contexts should be significantly higher than target sentences with inanimate subjects in non-fictional contexts.

Since the second conjunct forces personification, the acceptability of the target should thus be significantly lower than the controls in non-fictional contexts. In fictional contexts, it is possible to reinterpret a target sentence with an inanimate subject with the help of personification. Thus, we assumed the acceptability would increase compared to the target sentences in non-fictional contexts. Target sentences with an animate subject should be equally acceptable in both contexts. Thus, whereas the difference in acceptability between target sentences with animate subjects and inanimate subjects should be big in non-fictional contexts, it should be smaller in fictional ones.

Results

If a participant had to be excluded due to answering too many comprehension questions incorrectly (criterion for exclusion: correct answers < 70 %) we gathered additional data from other participants. In the end we thus did the analysis with data from 24 participants.

Participants' ratings were analyzed with two separate ANOVAs, one with an error term that was based on participant variability ($F1$) and one with an error term that was based on item variability ($F2$). The independent variables were SUBJECT (animate referent vs. inanimate referent) and CONTEXT (non-fictional vs. fictional).

Our predictions were borne out. We found a significant interaction between SUBJECT and CONTEXT [$F1(1,23) = 12.15, p < .01$; $F2(1,35) = 24.07, p < .001$]: Whereas there was no difference in acceptability of sentences with animate subjects between fictional and nonfictional contexts, the sentences with inanimate subjects were significantly more acceptable in fictional contexts than in nonfictional contexts. Moreover we found a main effect SUBJECT: Sentences with animate subjects were generally judged more acceptable [$F1(1,23) = 123.24, p < .001$; $F2(1,35) = 90.49, p < .001$]. We found also a main effect CONTEXT, which is due to the interaction, and not interpretable as a result: Fictional contexts were generally judged better than non-fictional contexts [$F1(1,23) = 6.11, p = .021$; $F2(1,35) = 15.90, p < .001$]. A t -test revealed that, still, the sentences with inanimate subjects in fictional contexts were not as acceptable as control sentences with animate subjects in fictional contexts [$t1(23) = 4.95, p < .001$; $t2(35) = 5.56, p < .001$]. The results are summarized in Figure 2 below.

Our findings suggest that the context influences the choice of adaptation mechanism. They show that fictional contexts license personification as expected. We find, however, that fictional contexts are not strong enough to make sentences with inanimate subjects in fictional contexts as acceptable as sentences with animate subjects. This implies that this adaptation mechanism is available, however not as straightforwardly as interpreting compositionally. We took our findings to pose perfect prerequisites for a reading time study.

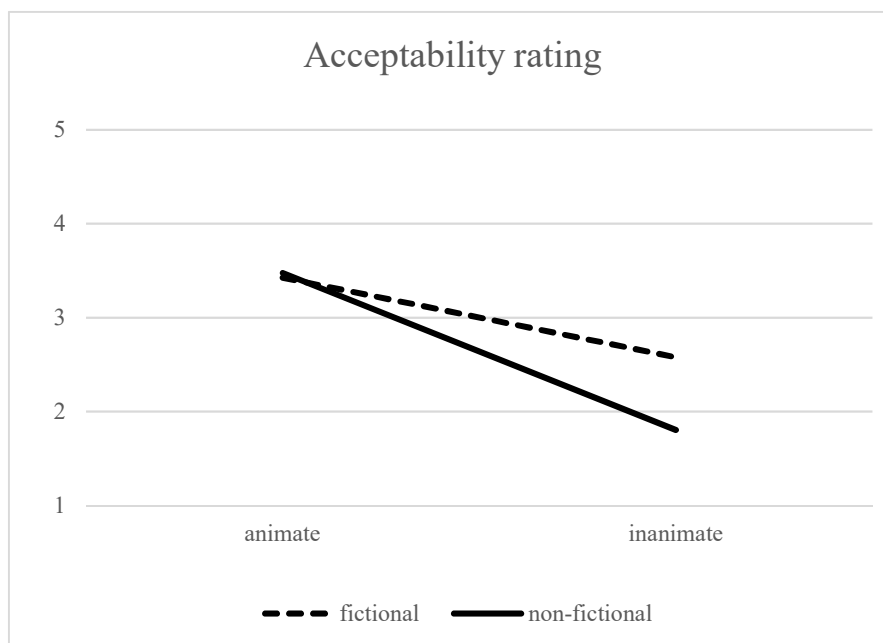


Figure 2. Average acceptability ratings for sentences with animate and inanimate subjects in fictional and non-fictional contexts

3.4 Participants

32 students from the University of Tübingen participated in the self-paced reading study and received a monetary reimbursement (7 Euros). They were all native speakers of German and were naïve to the purpose of the study.

3.5 Procedure

The experiment was run on a PC using E-Prime 2.0 software (Psychology Software Tools, Inc.). Participants were tested individually. The experiment was preceded by four practice trials. Sentences were presented to participants using a self-paced reading task with a moving window technique, see the regions indicated by the slashes in (12) and (13) above. Participants were instructed to read the context and the segments of the target sentence on their own natural pace. They pressed the space bar to begin the trial. After the first press of the space bar, the trial began with a display showing the context and the target sentence that were completely masked, i.e., each character of the context and the target sentence was represented by a dash. With the second press of the space bar, the whole context was shown by replacing the dashes with the corresponding characters. With the third press of the space bar, Region 1 of the target (= the subject) was shown by replacing the dashes with the corresponding characters while the context reverted to dashes. Any further press of the space bar led to the remasking of the current region and a concurrent demasking of the subsequent region.

In order to ensure that participants read the sentences carefully, one half of the trials ended with a *yes–no* comprehension question. Participants answered these question by pressing one of two designated keys.

3.6 Predictions

For the predictions, we consider each of the critical regions defined above separately.

The first region of interest is the intentional adverbial (Region 3) disclosing the semantic match / mismatch between the subject and the adverbial. We expected an interaction between SUBJECT and CONTEXT. Target sentences with an animate subject allow, irrespective of context, for a straightforward compositional interpretation, i.e., the context type should not affect the reading time on the adverbial in these cases. Target sentences with an inanimate subject require, irrespective of context, a meaning adaptation, i.e., the extra processing costs compared to the compositional interpretation should show up in increased reading times. If the distinction between interpolation and personification is cognitively real, the two adaptation mechanisms should lead to different processing costs, i.e., there should be a significant difference between target sentences with inanimate subjects in non-fictional contexts and fictional contexts. We expected that the destructive adaptation mechanism, i.e., personification, is reflected in longer reading times. Similar pattern is expected for the spill-over Region 4 (the PP).

The second region of interest is the second conjunct (Region 6). Again, we expected an interaction between SUBJECT and CONTEXT. Target sentences with an animate subject allow, irrespective of context, for a straightforward compositional interpretation of the second conjunct, i.e., the context type should not affect the reading time on the second conjunct in these cases. If the distinction between interpolation and personification is cognitively real, there should be significant differences in reading times on the second conjunct for target sentences with inanimate subjects depending on the context. If the fictional context licensed personification in the first conjunct, the second conjunct requiring a personification is in line with this interpretation. Thus, based on previous studies on personification, we expected that the reading times for target sentences with an inanimate subject in fictional contexts should not increase compared to the control condition. In contrast, if the non-fictional context only licensed interpolation in the first conjunct, the second conjunct requires a reanalysis. We expected the reading

times for target sentences with an inanimate subject in non-fictional contexts should increase significantly compared to the reading times in the fictional context.

3.7 Results

Participants that answered less than 70 % of the comprehension questions correctly were excluded from the analysis (four in total, i.e., we ran the analysis with 28 participants). Participants' reading times were analyzed for the Regions 3, 4 and 6. Analysis was done with linear mixed effect models using the R programming language and lmer function (Bates et al., 2015), with animacy and context as fixed and items/subjects as random factors. There were no effects on the adverbial (Region 3) or on the second conjunct (Region 6). There was a significant interaction on the PP (Region 4) ($SE = 0.4518$, $t = 2.886$, $p < .01$), but in the opposite direction of what was predicted, see Figure 3.

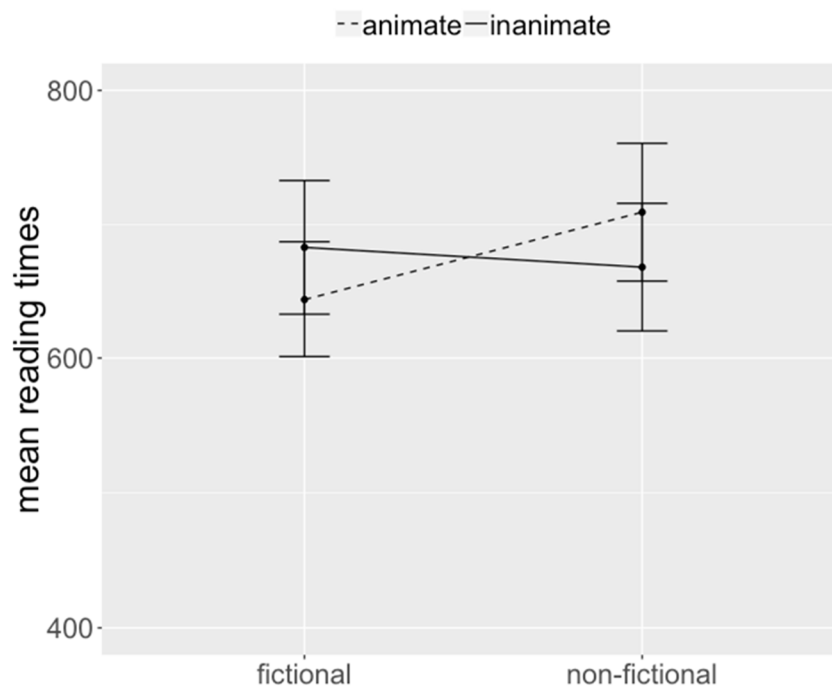


Figure 3. Average reading times (ms) on Region 4 (PP) in fictional and non-fictional contexts for animate and inanimate subjects

The sentences with animate subjects were affected by context, whereas sentences with inanimate subjects were not. Moreover, non-fictional contexts led to a significant increase of reading times of sentences with animate subjects.

3.8 Discussion

Our predictions for the reading time study were not borne out. We find unexpected effects in the control conditions with animate subject, where no difference between fictional and non-fictional contexts was predicted to occur due to the fact that a straightforward compositional interpretation was possible. Albeit the promising results of the acceptability rating study, the context manipulation did not seem to work properly for the reading time study. As opposed to our off-line data our on-line data do not provide any evidence for the two adaptation mechanisms we tried to compare to be cognitively real, or one being preferred over the other in a given context. Given the discrepancy between the findings we suspect the context manipulation to be too weak to evoke any effects during processing. We did some post-hoc analyses to exclude certain problems with the material. Due to the nature of our material there were some

items that more naturally fit a fictional or real context, i.e., reporting about Christmas morning might seem strange in a newspaper setting. We thus looked at the results per item. There was, however, no systematic difference between items where the expected pattern occurred and others. Furthermore, we looked at items with the most plausible items versus least plausible ones (i.e., the 18 items where the subjects got the highest plausibility ratings versus the others, see pretest I). There was no difference in effects, for neither group there was a significant interaction in the predicted direction. We also looked at the results by participants for the acceptability judgments to see whether there are participants who generally dislike fictional contexts. We suspected that those participants might display unexpected behaviour. However, for the acceptability judgments as well, we neither identified a group of participants disliking fictional contexts, nor a group of items which behaved unexpectedly. We thus assume that something else was at play during comprehension.

Our findings are surprising also and especially in view of previous evidence regarding reinterpretation mechanisms. To make sense of them we thus want to compare our settings with the ones used in previous studies to identify possible problems. We consider those studies which are most relevant to use, since they use fictionality as a factor.

Filik (2008) and Filik & Leuthold (2008) investigated world knowledge violations as in (21).

(21) The mouse picked up the dynamite.

They used fictionality as a factor in that they presented their items both in neutral and in ‘Tom & Jerry’ contexts. They found no extra processing costs if the animacy (including being capable of intentions) of the mouse was established by the context. One crucial difference between our study and theirs is that the nature of the violation is slightly different. Whereas ‘the mouse’ cannot pick up dynamite, it is already animate, and probably even attributed some human features. Furthermore it is not the predicate ‘picking up’ alone which causes the mismatch, it is the fact that it is dynamite which is surprising and requires revision. What has been looked at might thus rather be a violation of world knowledge (what can mice pick up?) rather than personifications of the mouse which require conceptual reinterpretation. The more interesting difference is that they investigated only one resolution mechanism, whereas we tried to compare two by using context manipulation. As a result, in our study participants might have been confused at some point as to what strategy is the right one. This might be due to the fact that the distinction between fiction and non-fiction became blurry throughout the experiment. The fact that we played with two reinterpretation mechanisms might thus explain our puzzling results. We would like to point out though that by using pseudo-randomization and a high number of filler items which tried to counterbalance certain preference we tried to avoid this effect. The fact that it still might have played a role hints at the power of fictional contexts and suggests that there is a non-local effect of having them in experimental studies.

Another prominent and relevant study in the literature on reinterpretation is Nieuwland & van Berkum’s (2006) experimental investigation of processing of animacy violations. They compared sentences such as the ones given in (22) and (23).

(22) The peanut was in love.

(23) The peanut was salted.

They included fictionality as a factor in that they presented both sentence types in a fantasy context. Similar to Filik et al.’s studies, they found no extra processing costs for (22), but rather for (23). Again, there was no comparison between two types of resolution strategies. Another crucial difference between their and our study is the nature of the context: Many more personifications were used in Nieuwland & van Berkum’s study. Our manipulation was restricted to the text type mentioned (fairy tale/newspaper), and two personifications. Moreover,

in Nieuwland & van Berkum's study, the target subjects were introduced as animate in the context. In contrast, we presented different items as animate in the context. The conflicting results raise the question whether we adapt our conceptual knowledge for every single entity at a time rather than generally accepting inanimate subjects as animate. It might be that the world knowledge is not adjusted immediately, as oftentimes claimed, but is already adjusted before the sentence is read in contexts where the subject was already presented as animate. Whereas context might thus overwrite lexico-semantic restrictions – as also suggested by our acceptability rating study – it seems to depend on the reliability and strength of the context. It definitely affects the immediacy of the adjustment process, our results suggest.

4 Summary and general discussion

Based on our findings, we conclude that fictional contexts license conceptual reinterpretation BUT they come with challenges for on-line studies. We did not see the same clear effects as found in previous studies. Upon closer inspection, we find that the crucial difference between these studies and ours regards the set-up of the context. We suspect that our contexts might have been too complex and too subtle at the same time, due to the fact that we wanted to keep fictional and non-fictional contexts maximally parallel. Making the context manipulation, i.e., fictionality, more transparent by using more personifications (as in Nieuwland & van Berkum, 2006) or by using familiar fictional contexts (as in Filik, 2008) seems to yield clearer results. This tells us, however, that making a context 'fictional', so that certain strategies are clearly licensed is a non-trivial exercise. The absence of an effect in previous studies suggests that no immediate adjustment process was needed anymore. Our aim was to see the processing cost of the reinterpretation mechanism during on-line comprehension. However, maybe the strategy we wanted to observe was not immediately available in the target sentence. Our findings suggest that we reluctantly adapt our conceptual knowledge for every single entity at a time, rather than willingly allow for this repair in a certain context type all of the time. Surely more research is needed to confirm this and understand better what the factors for immediate local adjustment are, and when and where context information plays a role in these adjustment mechanisms. Based on our findings, we want to tackle the following questions especially in follow-up studies:

First, what is an adequate way to establish and control that a context is *fictional* (and yet keep it maximally parallel to the non-fictional control)? One way to approach this is by introducing the subject which requires adjustment in the context. Second, how can we ensure that participants do not overgeneralize and that regular interpretation 'suffers' from the fact that fictionality becomes an option throughout? A first idea to pursue is to use context as a within-subjects factor. This would avoid any confusion between fictional and non-fictional contexts. Third, and relatedly, how can we ensure that fictional contexts are taken seriously enough in an experimental setting (to avoid the phenomenon of 'stops making sense')? One thought is that a global setting of a plausible fictional background (i.e., a fairytale setting with a queen and king and fantasy kingdom) might help participants to engage in the task without disregarding it as non-sense.

Ultimately, a lot more work needs to be done to understand what fictionality in a context does. Specifically, we need to understand under which circumstances fictional context allows for certain violations to be reinterpreted in a specific way to be measured online. Meanwhile, we believe that our work has important methodological implications and that using fictional contexts seems worthwhile in understanding better different strategies of reinterpretation and their processing costs.

Acknowledgments

The research reported here was supported by the German Research Foundation (DFG) as part of the Collaborative Research Center ‘The Construction of Meaning’ (SFB 833, projects A1 and A2). We would like to thank the members of projects A1 and A2, especially Matthias Bauer, Sigrid Beck, Sebastian Bücking and Claudia Maienborn, as well as the members of the Psycholinguistic Lunch of the SFB 833, and the participants of Linguistic Evidence for helpful feedback and discussion. Special thanks go to Saskia Brockmann for her help with preparing the material and running the experiments, as well as Joseph DeVeugh Geiss and Robin Hörnig for their invaluable input regarding our results and experimental set-up, as well as for their help with statistical analysis.

References

- Asher, N. (2011). *Lexical Meaning in Context: A Web of Words*. Cambridge, MA/New York City, NY: Cambridge University Press.
- Bade, N., & Beck, S. (2017). Lyrical texts as a data source for linguistics. *Linguistische Berichte*, 251, 317-356.
- Baggio, G., Choma, T., van Lambalgen, M., & Hagoort, P. (2009). Coercion and compositionality. *Journal of Cognitive Neuroscience*, 22(9), 2131-2140.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1-48.
- Bauer, M., Bade, N., Beck, S., Dörge, C., von Eckartsberg, B., Niefer, J., Ottschofski, S., & Zirker, A. (2015). Emily Dickinson’s ‘My life had stood a loaded gun’ – An interdisciplinary analysis. *Journal of Literary Semantics*, 44(2), 115-40.
- Blutner, R. (1998). Lexical pragmatics. *Journal of Semantics*, 15, 115-162.
- Bott, O. (2010). *The Processing of Events*. Amsterdam: John Benjamins.
- Brennan, J., & Pyllkänen, L. (2008). Processing events: Behavioral and neuromagnetic correlates of aspectual coercion. *Brain and Language*, 105, 132-143.
- Buscher, F. (2013). Im Spannungsfeld von Semantik und Pragmatik: Zur Bedeutungskonstitution von Einstellungsadverbialen. *Zeitschrift für Sprachwissenschaft*, 32(2), 135-179.
- Buscher, F. (2018). *Kompositionalität und ihre Freiräume: Zur flexiblen Interpretation von Einstellungsadverbialen*. Tübingen: Stauffenburg.
- de Swart, H. (2011). Mismatches and coercion. In C. Maienborn, K. von Heusinger & P. Portner (Eds.), *Semantics. An International Handbook of Natural Language Meaning, Vol. 1* (pp. 574-597). Berlin/New York City, NY: de Gruyter.
- Dickinson, E. (1955). *The Poems of Emily Dickinson*. Edited by Thomas H. Johnson. Cambridge, MA: Belknap Press.
- Dölling, J. (2005). Semantische Form und pragmatische Anreicherung: Situationsausdrücke in der Äußerungsinterpretation. *Zeitschrift für Sprachwissenschaft*, 24, 159-225.
- Dowty, D. (2007). Compositionality as an empirical problem. In Ch. Barker & P. Jacobson (Eds.), *Direct Compositionality* (pp. 23-101). Oxford: Oxford University Press.

- Eckardt, R. (2003). Manner adverbs and information structure: Evidence from the adverbial modification of verbs of creation. In E. Lang, C. Maienborn & C. Fabricius-Hansen (Eds.), *Modifying Adjuncts* (pp. 261-305). Berlin: de Gruyter.
- Egg, M. (2005). *Flexible Semantics for Reinterpretation Phenomena*. Stanford: CSLI Publications.
- Filik, R. (2008). Contextual override of pragmatic anomalies: Evidence from eye movements. *Cognition*, *106*, 1038-1046.
- Filik, R., & Leuthold, H. (2008). Processing local pragmatic anomalies in fictional contexts: Evidence from the N400. *Psychophysiology*, *45*, 554-558.
- Frisson, St., & McElree, B. (2008). Complement coercion is not modulated by competition: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(1), 1-11.
- Kuperberg, G. R., Choi, A., Cohn, N., Paczynski, M., & Jackendoff, R. (2010). Electrophysiological correlates of complement coercion. *Journal of Cognitive Neuroscience*, *22*(12), 2685-2701.
- McElree, B., Pylkkänen, L., Pickering, M. J., & Traxler, M. J. (2006). A time course analysis of enriched composition. *Psychonomic Bulletin & Review*, *13*(1), 53-59.
- McElree, B., Traxler, M. J., Pickering, M. J., Seely, R. E., & Jackendoff, R. (2001). Reading time evidence for enriched composition. *Cognition*, *78*, B17-B25.
- Nieuwland, M. S., & van Berkum, J. J. A. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience*, *18*, 1098-1111.
- Pagin, P., & Westerstähl, D. (2011). Compositionality. In C. Maienborn, K. von Heusinger & P. Portner (Eds.), *Semantics: An International Handbook of Natural Language Meaning, Vol. 1* (pp. 96-123). Berlin/New York City, NY: de Gruyter.
- Piñango, M. M., Winnick, A., Ullah, R., & Zurif, E. (2006). Time-course of semantic composition: The case of aspectual coercion. *Journal of Psycholinguistic Research*, *35*, 233-244.
- Piñango, M. M., Zurif, E., & Jackendoff, R. (1999). Real-time processing implications of enriched composition at the syntax-semantics interface. *Journal of Psycholinguistic Research*, *28*, 395-414.
- Pittner, K. (2004). Where syntax and semantics meet: Adverbial positions in the German middle field. In J. R. Austin, S. Engelberg & G. Rauh (Eds.), *Adverbials: The Interplay between Meaning, Context, and Syntactic Structure* (pp. 253-287). Amsterdam: John Benjamins.
- Pustejovsky, J. (2011). Coercion in a general theory of argument selection. *Linguistics*, *49*(6), 1401-1431.
- Pylkkänen, L., Martin, A. E., McElree, B., & Smart, A. (2009). The anterior midline field: Coercion or decision making? *Brain & Language*, *108*, 184-190.
- Pylkkänen, L., & McElree, B. (2007). An MEG study of silent meaning. *Journal of Cognitive Neuroscience*, *19*(11), 1905-1921.
- Scheepers, Ch., Keller, F., & Lapata, M. (2008). Evidence for serial coercion: A time course analysis using the visual-world paradigm. *Cognitive Psychology*, *56*, 1-29.

- Schumacher, P. B. (2011). 'The hepatitis called ...': Electrophysiological evidence for enriched composition. In J. Meibauer & M. Steinbach (Eds.), *Experimental Pragmatics/Semantics* (pp. 199-219). Amsterdam/Philadelphia, PA: John Benjamins.
- Schumacher, P. B. (2013). When combinatorial processing results in reconceptualization: Toward a new approach of compositionality. *Frontiers in Psychology*, 4, 677.
- Schumacher, P. B. (2014). Content and context in incremental processing: 'The ham sandwich' revisited. *Philosophical Studies*, 168(1), 151-165.
- Schumacher, P. B., & Weiland, H. (2011). Reading Brecht and talking to the espresso: Electrophysiological investigations of conventional and novel metonymy. In A. Lenci, S. Padó, J. Utt & A. Zarcone (Eds.), *Proceedings of Metonymy 2011* (pp. 23-28). Stuttgart.
- Todorova, M., Straub, K., Badecker, W., & Frank, R. (2000). Aspectual coercion and the online computation of sentential aspect. In L. R. Gleitman & A. K. Joshi (Eds.), *Proceedings of the Twenty-Second Annual Conference of the Cognitive Science Society* (pp. 1-6). Philadelphia, PA: University of Pennsylvania.
- Traxler, M. J., McElree, B., Williams, R. S., & Pickering, M. J. (2005). Context effects in coercion: Evidence from eye movements. *Journal of Memory and Language*, 53, 1-25.
- Traxler, M. J., Pickering, M. J., & McElree, B. (2002). Coercion in sentence processing: Evidence from eye-movements and self-paced reading. *Journal of Memory and Language*, 47, 530-547.
- Weiland, H., Bambini, V., & Schumacher, P. B. (2014). The role of literal meaning in figurative language comprehension: Evidence from masked priming ERP. *Frontiers in Human Neuroscience*, 8, 583.