Literary Psycholinguistics and the Poem

Inauguraldissertation
zur Erlangung des akademischen Grades
eines Dr. Phil.,
vorgelegt dem Fachbereich 05 – Philosophie und Philologie
der Johannes Gutenberg-Universität
Mainz

von
Stefan Blohm
aus Erlenbach am Main
Mainz
2020

To You.

Contents

Chapter 1 has been published in almost identical form as:
Blohm, S., Menninghaus, W., & Schlesewsky, M. (2017).
Sentence-level effects of literary genre: Behavioral and electrophysiological evidence.
Frontiers in Psychology, 8(1887). doi: 10.3389/fpsyg.2017.01887
Please refer to the original publication when citing.

Author contributions: SB, WM, and MS developed the hypotheses. SB and MS developed the study design. SB developed the materials, collected the data and performed the data analysis. SB, WM, and MS interpreted the data. SB drafted the manuscript, and MS and WM provided critical revisions.

Chapter 2 has been submitted for publication in slightly different form as:
Blohm, S., Menninghaus, W., Schlesewsky, M., & Scharinger M.
Text type attribution modulates pre-stimulus alpha power in sentence reading.

Author contributions: SB, and MScha developed the hypotheses, performed the data analysis, interpreted the data, and drafted the manuscript. MSchl and WM provided critical revisions.

Chapter 3 has been submitted for publication in slightly different form as:
Blohm, S., Versace, S., Methner, S., Wagner, V., Schlesewsky, M., & Menninghaus, W.
Behavioral differences between poetry and prose reading: Evidence from eye tracking and speech recordings.

Author contributions: SB and SV developed the hypotheses and the study design. SB, SV and SM developed the materials and collected the data. SB, SV and VW performed the data analysis and interpreted the data. SB drafted the manuscript. SV, SM, VW, WM and MS provided critical revisions.

Chapter 4 has been published in almost identical form as:
Blohm, S., Wagner, V., Schlesewsky, M., & Menninghaus, W. (2018).
Sentence judgments and the grammar of poetry: Linking linguistic structure and poetic effect.
Poetics, 69, 41-56. doi: 10.1016/j.poetic.2018.04.005
Please refer to the original publication when citing.

Author contributions: SB developed the hypotheses, the study design and the materials, and collected the data. SB performed the data analysis of Exp. 3a; SB and VW performed the data analysis of Exp. 3b. SB, VW, MS and WM interpreted the data. SB drafted the manuscript, and VW, MS and WM provided critical revisions.

List of figures

List of tables

Language is not a thing which leads a life of its own outside of and above human beings, [...] it has its true existence only in the individual.

Osthoff & Brugmann (1878)

0 Introduction

As Louise Rosenblatt might have said, but in fact didn't, poetry is not a thing which leads a life of its own outside of and above human beings—it has its true existence only in the individual. The present thesis investigates some of the implications of this truism, tapping into different aspects and stages of the literary reception process by means of behavioral and electrophysiological measures. The reported experiments represent a selection of the research I was allowed to conduct at the MPI for Empirical Aesthetics, Frankfurt (Germany). Although they make use of various methods to focus on distinct aspects of literary reading, they all share a common goal, which is to investigate how readers' conceptions of the poem influence the processing and evaluation of verbal stimuli.

From a linguist's perspective, these studies add to our understanding of context-dependent language processing and thus fall into the domain of experimental pragmatics. From a literary scholar's point of view, they contribute to a cognitive theory of literary genres. Although these intersecting interests render the present thesis a cross-disciplinary endeavour, the adopted view represents a linguistic perspective on literary texts, and the adopted multi-method approach borrows its paradigms and protocols from experimental psycholinguistics. In this sense the present thesis illustrates what might be alternatively seen as a

potential branch of psycholinguistics or a core area in the empirical study of literature: *Literary Psycholinguistics*—the systematic study of the processes and mechanisms involved in comprehending literary texts and in perceiving and evaluating their characteristic linguistic features.

Literary Psycholinguistics thus defined primarily aims to relate what Marr (1982) termed the *computational* level of analysis in the study of information-processing systems to the *algorithmic* level of analysing such systems. The former is dominant in literary scholarship, the latter still prevalent in psycholinguistics despite increasingly successful efforts to integrate the insights gleaned in decades of empirical research with the known neurobiological principles of the human brain. This focus on the computation–algorithm nexus is most evident here in the studies that seek to connect evaluations of literary stimuli to distinct stages of stimulus processing; only one study attempts to relate the computational to the *implementational* level of analysis.

Genre conceptions are schematic

Since the computational problem considered here is genre-specific literary comprehension, we must first ask ourselves: What exactly is a ”literary genre”? To date, there is no consensus among literary scholars regarding the definition of this crucial notion. While the term traditionally denotes the basic modes of literary art (i.e., the *lyric*, *epic*, and *dramatic*) or their subcategories (e.g., the novella, the ballad), it is also used to distinguish the broadest categories of composition like *poetry* or *prose* (Baldick, 2015). Here, we will use the term in the latter sense—following Steen's (1999) cognitive theory of literary genres—and mean it to include 'a poem' and in one instance even 'literary prose'. Steen (1999: 112) argues that "it is presumably the level of genre that embodies the basic level concepts, whereas subgenres are the conceptual subordinates, and more abstract classes of discourse are the superordinates". Hence, 'the poem' will be understood as a basic level in a hierarchical and prototype-theoretical conception of how readers represent literary categories. Literature, according to this view, constitutes the more abstract class of discourse, i.e., the conceptual

superordinate. In representational terms, the levels of this hierarchy of literary text types correspond to *memory schemata* that connect genre-related knowledge from different domains. The notion of the schema seems to have been introduced by Kant (1781), but it is Bartlett (1932) who is usually credited with demonstrating the importance of schemata for memory retrieval during comprehension. As Kant (1781: 182-183) originally observed in his *Critique of Pure Reason*:

> The *schema* of a concept [...] signifies a rule according to which my imagination can delineate the figure [...] in a general manner, without limitation to any single determinate figure such as experience, or any possible image that I can represent *in concreto*, actually presents.

More recent work has elaborated this Kantian concept. Schemata and the related constructs of *frames* (Minsky, 1975) and *scripts* (Schank & Abelson, 1975, 1977) model the conceptual architecture of man and thus help to understand human comprehension. As Rumelhart and Ortony (1977: 101) put it:

> Schemata are data structures for representing the generic concepts stored in memory. They exist for generalized concepts underlying objects, situations, events, sequences of events, actions, and sequences of actions. Schemata are not atomic. A schema contains, as part of its specification, the network of interrelations that is believed to generally hold among the constituents of the concept in question. Schemata, in some sense, represent stereotypes of these concepts.

Based on this conception, they propose four basic properties of memory schemata that are compatible with the cognitive theory of genres outlined by Steen (1999):

1. Schemata contain variables
2. Schemata embed, one within the other
3. Schemata exist at all levels of abstraction
4. Schemata represent knowledge, not definitions

The Kantian notion of the general outline that lacks concrete realizations captures the opposition between literary genre and individual text. Applying an elaborated schema view along the lines of Rumelhart and Ortony to literary genres (see Ryan, 1979; Verdaasdonk, 1982; Viehoff, 1995 for similar proposals) admits for intra-generic variation between texts (cf. 1.), reflects the assumed hierarchical organization of literary genre conceptions (cf. 2.), captures intertextual similarities with respect to form and content (cf. 3.), and parsimoniously accounts for the fact that literary genres are fuzzy categories (cf. 4.). While schema-based approaches to literature have—at least in the context of story comprehension—also been criticised for being too loosely defined to be of practical value (e.g., Thorndyke & Yekovich, 1980), their theoretical usefulness and refinement will remain a matter of debate as long as their hypothetical effects on literary comprehension are not subject to systematic investigation.

Reading strategies are genre-specific

Supporting evidence for the influence of schematic genre conceptions on literary comprehension stems from contrastive studies that compared literary and non-literary reading modes (e.g., Hanauer, 1998; Hoffstaedter, 1987; Meutsch, 1987; Peskin, 2010; Zwaan, 1991, 1993, 1994). A theoretical account of such findings was proposed by Zwaan (1993, 1996) in his outlined model of literary comprehension.

Zwaan's model is an extension of the influential *Construction-Integration (CI) Model* (Kintsch, 1988, 1998), from which it inherits the system of hierarchical mental text representation. Both models assume that the linguistic *surface form* of a text is incrementally converted into an abstract and amodal *propositional textbase* which, in turn, feeds into mental *situation models* that integrate the propositional content with comprehenders' world knowledge. But whereas the original CI model conceives situation models as enriched and edited propositional networks, Zwaan (1996) assumes that they are representationally distinct from the textbase and may include visuo-spatial information (see Ingarden (1931) for an earlier, very similar conception of the literary text). Zwaan's model

further adopts the notion of the *cognitive control system* from the CI model, i.e., the idea that experienced readers have developed discourse-specific reading strategies that regulate the process and outcome of text comprehension. For instance, the *news-comprehension control system* (van Dijk & Kintsch, 1983) aims at updating the reader's world knowledge and assumes that texts are composed and structured to provide concise and accurate instructions for constructing coherent textbases and situation models (whereas linguistic form is a mere vehicle). The *literary-comprehension control system* (LCCS)—the major theoretical innovation of Zwaan's model—enables comprehenders to deal with the "deliberate inconsiderateness" of literary texts by adjusting reading behavior, interpretive strategies, as well as the allocation of attention and processing resources more generally. In support of his proposal, Zwaan (1991, 1993, 1994) was able to demonstrate that the process and outcome of literary comprehension differ from those of news comprehension independently of textual variables, which, he argued, reflects the (potential) indeterminacy of literary texts. Specifically, the LCCS reduced reading speed and apparently downgraded the importance of situational representations while promoting the retention of surface form and the construction of propositional representations.

The present thesis

Building on Zwaan's comparative results, the present research focuses on linguistic (=surface from) and semantic (=propositional) processing and evaluation, comprising investigations of genre-typical linguistic knowledge and text performance as well as genre-appropriate standards of evaluation. All of the experiments presented here share the idea that something akin to an LCCS modulates text comprehension. But rather than seeing the LCCS as a single system, I will present evidence that favors a conception of the LCCS as a set of distinct genre-appropriate reading modes. According to this conception, "literariness" is categorial rather than continuous.

Connecting Steen's prototype-theoretic conception of genres to Zwaan's notion of the literary-comprehension control system, I derive the general hypothesis that the poetry schema—the basic literary category that is arguably most dissimilar to everyday communication—can be activated by mere external categorization, and that activating the poetry schema triggers genre-appropriate processing strategies and standards of evaluation that are distinct from other literary and non-literary modes of comprehension. This claim is neither new nor very bold, and I believe that I spoil little if I hint that our results indicate that it is correct. But what exactly this claim entails, which aspects of language comprehension *are* affected, and *how*, or how subjectively perceived qualia of texts relate to specific linguistic features remains—due to the sparsity of empirical evidence—far from clear. It is this evidential gap that the present thesis helps to close.

Chapters 1–4 report a series of empirical studies in the format of self-contained articles, which they are. These chapters can be read independently; in the one instance where previously presented information is indispensible for comprehension (Chapter 2), reference is made to the section that supplies the crucial information. Chapter 5 summarizes the results and offers some conclusions regarding poetry-specific reading and literary comprehension more generally.

Chapter 1 reports two experiments that employed sentence judgments (*Experiment 1a*[1]) and event-related brain potentials (*Experiment 1b*) to investigate the influence of genre categorization (poetry vs. no categorization) on the evaluation and the online comprehension of single sentences. Specifically, we examined whether external genre categorization modulates the use of prosodic information in sentence reading and whether it affects the processing and the perceived meaningfulness of semantically (in)congruent utterances. The use of single sentences rather than entire texts allowed us to isolate *a priori* effects of genre categorization, i.e., processing adjustments triggered by the category and not by the (con)text.

[1]The instruction texts and stimulus materials of all experiments are deposited at the *Open Science Framework* and can be accessed at https://osf.io/ht6ar/.

Chapter 2 presents additional analyses of the EEG data collected in *Experiment 1b*. Aiming to pin down effects of genre categorization on anticipatory attentional states, these analyses (*Experiment 1c*) focused on pre-stimulus alpha power as an index of selective attention prior to the first linguistic input as well as on further early (<350ms) ERP effects of genre categorization.

Chapter 3 reports a study that combined eye tracking and speech recordings (*Experiment 2*) to compare distinct literary comprehension modes (literary prose vs. poetry) during oral text reading. Moving beyond aprioristic effects of genre categorization, this study specifically aimed to reveal genre-specific dynamics by studying the interaction of text category and (con)text during literary reading.

Chapter 4 presents two experiments that used sentence judgments to examine the influence of grammatical and prosodic variation on the grammatical and literary-aesthetic evaluation of poetic verse (*Experiment 3a*) and of regular sentences (*Experiment 3b*). These studies aimed to demonstrate that experienced readers systematically associate the genre of poetry with selected grammatical structures and features.

Chapter 5 summarizes the results of the presented experiments, relates them to distinct stages of the reception process and to Zwaan's theoretical model of literary comprehension, discusses limitations of the present work and identifies possible directions for future research in literary psycholinguistics.

The notion 'text type' is defined as: a set of stored instructions for gathering and processing the discernible, multi-level features of texts.

Robert de Beaugrande (1978: 3)

1

Poetry-specific sentence processing A: Behavioral and ERP effects

1.1 Poetic license and genre-specific reading

We use language to express ourselves, to refer to the external world, and to appeal to others (Bühler, 1934). Consequently, the selection and combination of speech sounds and meaningful elements serves to convey messages efficiently. Additional constraints on linguistic form derive from further communicative functions (Jakobson, 1960). For instance, in some types of non-spontaneous discourse – such as literature, rhetorical speeches, advertising, and other forms of propaganda – phonological form is consciously manipulated for a variety of purposes and effects (e.g., McGlone and Tofighbakhsh, 2000; Knoop et al., 2016; Menninghaus, Bohrn, Altmann, Lubrich & Jacobs, 2014; Menninghaus, et al., 2015; Wimsatt, 1944).

Poetry is particularly interesting in this respect: as the writing process is not subject to the same practical constraints as daily communication, poets are free to create texts that satisfy a range of additional self-imposed formal constraints on the selection and combination of linguistic elements (e.g., Kintsch, 1998; Leech, 1969; Levin, 1962). Poetic meter, for instance, typically constrains the number of phonological elements (e.g., syllables or morae) that may occur in a line of verse, as well as the distribution of prominence features (accent, length, tone) across these elements (e.g., Fabb, 2015). Across centuries and literary traditions, poets have taken the liberty of deviating from the rules of a given language in order to meet self-imposed constraints, and of using newly coined, archaic, or otherwise deviant lexical and grammatical forms (Aristotle, 1961; Leech, 1969; Youmans, 1983; Rice, 1997; Sanni, 1993).

The optional presence/absence of (historical) affixes, for example, provides creative options for altering the number of (unstressed) syllables. Some poetic traditions make extensive use of such options, particularly in languages that diachronically reduce complexity in their inflectional systems, e.g., English or German (Leech, 1969; Rice, 1997). Semantics in poetry enjoys similar licenses. Poets assert that the language of poetry is “vitally metaphorical” (Shelley, 1840); some theoreticians claim that figurative speech and semantic polyvalence are the defining features of poetic texts (e.g., Thorne, 1988). As a result of these for-

mal and semantic licenses, poetic discourse differs significantly from everyday verbal behavior (e.g., Wordsworth, 1802). However, sender and receiver typically convene in their conception of the type of discourse they are engaged in and, consequently, both sides know how to act and what to expect (Beaugrande, 1978; Hanauer, 1998a, 1998b; Kintsch, 1998; van Dijk & Kintsch, 1983). Genre awareness influences how we articulate a text (e.g., Barney, 1999; Byers, 1979), or how quickly we read it and what we recognize or recall of its surface structure and content (Fischer, Carminati, Stabler & Roberts, 2003; Hanauer, 1998a, 1998b; Tillmann & Dowling, 2007; Zwaan, 1991, 1994); the clearest evidence for genre-specific reading comes from studies that reported effects of genre attribution on the processing of identical texts (e.g., Schumacher & Avrutin, 2011; Zwaan, 1991, 1994).

But what is specific about poetry? Beaugrande (1978: 24) proposed that "readers have as part of the text-type frame the instructions to attend to sound recurrences in the assumption that these are not random". Previous research on the processing of poetry has confirmed that sound recurrences influence how verbal stimuli are processed (Chen et al., 2016; Hoorn, 1996; Menninghaus et al., 2014; Obermeier et al., 2015; Vaughan-Evans et al., 2016) and evaluated (Kraxenberger & Menninghaus, 2016; Menninghaus et al., 2014; Menninghaus, Wagner, Wassiliwizky, Jacobsen & Knoop, 2017; Obermeier et al., 2013; van Peer, 1990). It can, however, not be taken for granted that these results indeed reflect adapted processing routines and increased attention to phonological structure due to genre awareness.

1.1.1 The present study: rationale, design and materials

To address this question, the present study contrasted the processing of prosodic variation in identical sentence contexts with and without explicit genre attribution. We manipulated sentence prosody by means of a slightly archaic case suffix whose optional presence or absence constitutes one of German poetry's routine licenses for balancing metrical structure. In addition, we manipulated the semantic congruency of the sentences to test whether the poetry mode of

reading immediately affects the computation of meaning or whether semantic effects are to be found only at later stages of comprehension.

As previous studies on genre-specific reading have typically used global measures such as reading rate (Hanauer, 1998a; Zwaan, 1991, 1994), or investigated sentence-level categories such as newspaper headlines (Schumacher & Avrutin, 2011), it is still uncertain whether effects of literary genres are already detectable in the online processing of single sentences. Therefore, a main objective of the present study was to provide empirical evidence for genre-specific reading at the sentence level. To this end, we analyzed event-related brain potentials (ERPs) in the human electroencephalogram (EEG) recorded during sentence reading.

Note that while we are testing for genre-dependent adaptations that are aprioristic, we neither imply that adaptations necessarily precede processing (although categorization usually does), nor do we claim that they are exclusively genre-dependent. Rather, we assume that the schematic text type representations that experienced readers have acquired are associated with specific processing defaults (Fischer et al., 2003; Hanauer, 1998a; Schumacher & Avrutin, 2011; Zwaan, 1991, 1994), and that the textual variables of actual discourse may override these defaults (e.g., for semantic processing, see Nieuwland and Van Berkum, 2006). If this assumption is correct, and if we are correct in assuming that attention to sound recurrences and expectation of unusual semantic combinations are part of the poetic reading mode (Beaugrande, 1978; Gibbs, Kushner & Mills, 1991; Thorne, 1988), then we should observe genre-dependent differences in the offline evaluation (Exp. 1a) and online processing (Exp. 1b) of verbal stimuli already at the level of the single sentence. If, on the other hand, genre-dependent adaptations are triggered by textual cues (e.g., graphic layout, prototypical structure, sound patterning), then the predicted effects should be independent of the genre.

A secondary goal of the present study was to investigate potential effects of genre awareness on memory encoding by providing behavioral evidence from a probe recognition task: at the end of each trial, participants were presented with a word and indicated per button press whether or not it had occurred in

the preceding sentence. Not only did this task ensure attentive reading, it can also yield information about the surface structure representation of the sentence after sentence wrap-up. In previous research, delayed recognition and recall tasks revealed enhanced memory for the surface structure of literary versus non-literary texts (Hanauer, 1998a, Zwaan, 1991, 1994) and for poetry versus prose (Hanauer, 1998a, 1998b; Tillmann & Dowling, 2007). Tillmann and Dowling (2007:636) argue that this effect is due to the systematic sound patterning that results in "the creation of richer and more precise memory traces", and thus in more effective retrieval. While this interpretation is plausible, it is also possible that the poetic memory effect partially reflects differential memory encoding during processing of a given perceptual chunk. Thus, if we find a genre effect on recognition accuracy in the absence of the intricate sound patterning of a whole poem, this would constitute strong evidence for the idea that the poetic memory effect partially depends on an aprioristic modulation of the level of processing (e.g., Craik & Lockhart, 1972; Craik & Tulving, 1975).

Design

These questions were addressed with a mixed design: instruction was manipulated between participants, but sentence prosody and semantic congruency were crossed within participants. Participants were assigned to one of two instruction groups: Instruction was either neutral with respect to text type ($\text{Neutral}_{\text{INST}}$), or it identified the stimuli as verses, i.e., as single lines of poetry ($\text{Poetry}_{\text{INST}}$). Sentences differed in terms of rhythmical regularity induced by alternative Dative case markers (Zero_{DAT} = irregular; $\text{Schwa}_{\text{DAT}}$ = regular) and Semantic congruency ($\text{Congruent}_{\text{SEM}}$ vs. $\text{Incongruent}_{\text{SEM}}$). Note that both rhythm patterns occur naturally in everyday speech, but that only the $\text{Schwa}_{\text{DAT}}$ pattern features strict rhythmical alternation (cf. Table 1.1).

Materials

We constructed 48 sets of sentences exemplified in Table 1.1; the full set of stimulus materials is available at https://osf.io/ht6ar/. Sentences began with a prepo-

sitional phrase (PP) followed by a verb (V) and the subject noun phrase (NP). Four sentence variants per item set crossed Dative marking in the sentence-initial PP ($Zero_{DAT}$ vs. $Schwa_{DAT}$) and Semantic congruency of the subject NP ($Congruent_{SEM}$ vs. $Incongruent_{SEM}$). For the EEG study, we inserted a postcritical adverbial (ADV) of one to three words after the critical NP position in order to avoid confounding effects of sentence wrap-up (Friederici & Frisch, 2000).

This configuration allowed us to use the initial PP as a prosodic prime that established an alternating rhythm pattern (trochaic or iambic)[1] . Thus, we modulated prosodic expectations for the critical verb in line with the Principle of Rhythmic Alternation, a preference for, and a tendency towards, the harmonious alternation of stressed and unstressed syllables (Jespersen, 1933; Selkirk, 1984; see e.g., Bohn, Knaus, Wiese & Domahs, 2013; Henrich, Alter, Wiese & Domahs, 2014 for online evidence). This sentence structure further ensured that verb information was already available when the subject NP was encountered, which allowed us to manipulate semantic congruency by varying the semantic fit between verb and subject.

Consistent with the practice of poetic license in lyrical poetry we used historical vs. contemporary case suffixes in sentence-initial constituents to manipulate word prosody without altering syntax or semantics. This morpho-phonological manipulation exploited a peculiarity of German grammar, the *schwa/zero alternation* (Szczepaniak, 2010; Wiese & Speyer, 2015): Two dative[2] case suffixes coexist in the inflectional paradigm of strong masculine and neuter nouns, one realized as a reduced vowel ($Schwa_{DAT}$), the other without overt marking on the noun ($Zero_{DAT}$). Synchronically, $Zero_{DAT}$ marking constitutes the default form, being far more frequent and preferred in most registers of contemporary spoken and written German (Szczepaniak, 2010). However, the loss of the overt case suffix is a recent, ongoing development, so that the $Schwa_{DAT}$ suffix has not been fully replaced. It is still used in formal and literary written registers, in non-casual speech, and in idiomatic expressions, and thus sufficiently famil-

[1] Trochaic = stressed-unstressed; iambic = unstressed-stressed.

[2] Dative case prototypically marks recipients in transfer-of-possession events, as in: '*Pat gave the book to John*'. A range of prepositions also assign dative case, e.g. those expressing locations: [*auf* [dem_{DAT} $Dach_{DAT}$]$_{NP}$]$_{PP}$ = [*on* [*the roof*]$_{NP}$]$_{PP}$

Table 1.1: Example materials. Example sentences and English translations for the four within-participant conditions; boldface indicates word stress; critical positions are underlined.

Condition (within participants)	Example
$\text{Schwa}_{\text{DAT}}/\text{Congruent}_{\text{SEM}}$	[In **die**sem **Bet**te]$_{\text{PP}}$ [**schnarcht**]$_{\text{verb}}$ [das **Mä**dchen]$_{\text{NP}}$ [**laut**]$_{\text{ADV}}$. lit: [In this$_{\text{DAT}}$ bed$_{\text{DAT}}$] [snores] [the$_{\text{NOM}}$ girl$_{\text{NOM}}$] [loudly] *In this bed, the girl snores loudly.*
$\text{Schwa}_{\text{DAT}}/\text{Incongruent}_{\text{SEM}}$	[In **die**sem **Bet**te]$_{\text{PP}}$ [**schnarcht**]$_{\text{verb}}$ [das **Las**ter]$_{\text{NP}}$ [**laut**]$_{\text{ADV}}$. lit: [[In this$_{\text{DAT}}$ bed$_{\text{DAT}}$] [snores] [the$_{\text{NOM}}$ vice$_{\text{NOM}}$] [loudly] *In this bed, vice snores loudly.*
$\text{Zero}_{\text{DAT}}/\text{Congruent}_{\text{SEM}}$	[In **die**sem **Bett**]$_{\text{PP}}$ [**schnarcht**]$_{\text{verb}}$ [das **Mä**dchen]$_{\text{NP}}$ [**laut**]$_{\text{ADV}}$. lit: [In this$_{\text{DAT}}$ bed$_{\text{DAT}}$] [snores] [the$_{\text{NOM}}$ girl$_{\text{NOM}}$] [loudly] *In this bed, the girl snores loudly.*
$\text{Zero}_{\text{DAT}}/\text{Incongruent}_{\text{SEM}}$	[In **die**sem **Bett**]$_{\text{PP}}$ [**schnarcht**]$_{\text{verb}}$ [das **Las**ter]$_{\text{NP}}$ [**laut**]$_{\text{ADV}}$. lit: [In this$_{\text{DAT}}$ bed$_{\text{DAT}}$] [snores] [the$_{\text{NOM}}$ vice$_{\text{NOM}}$] [loudly] *In this bed, vice snores loudly.*

iar to any proficient speaker of German. Importantly, these word forms do not constitute violations; they are perceived as well-formed but stylistically marked, i.e., register-dependent, variants. According to a recent suggestion (Wiese & Speyer, 2015:526), this alternation "exists even in the standard language" and "may exist in the same register of the language without any semantic or grammatical difference between the two forms".

While functionally equivalent in terms of grammar, these suffixes differ in terms of their effect on word prosodic structure: Zero_{DAT} marking does not affect the monosyllabic nouns we used, but $\text{Schwa}_{\text{DAT}}$ marking adds a reduced syllable nucleus and results in trochaic word forms (cf. Supplementary material). This word-prosodic difference motivates the seemingly free variation of both forms in German lyrical poetry, and it has been identified as one of the factors that co-determine their variation even in contemporary standard German (Kentner, 2015; Wiese, 2000, 2016; Wiese & Speyer, 2015). An example comparison of the two suffixes is provided in Table 1.2.

Table 1.2: Comparison of the two dative suffixes.

	Example	Stylistically marked	Number of syllables	Syllable structure	Stress on final syllable
$Zero_{DAT}$:	Tod (*death*$_{DAT}$)	no	1	CVC	yes
$Schwa_{DAT}$:	Tode (*death*$_{DAT}$)	yes	2	CV.CV	no

We used a variety of prepositions and modifiers to vary the form of sentence-initial PPs while keeping their basic constituent structure identical: [preposition – [$modifier_{DAT}$ – $noun_{DAT}$]$_{NP}$]$_{PP}$. All of the prepositions we used assign dative case to PP-internal noun phrases, and we carefully chose unambiguously case-marked modifiers (determiners, demonstratives or adjectives) to exclude confounding effects of case ambiguities.

Despite their length, we decided to present entire PPs at a time so they could serve as efficient prosodic primes. This presentation mode allows readers to immediately assign phrasal accent to the modified nouns in the PP, including monosyllables in the $Zero_{DAT}$ conditions. Thus, we ensured that modified nouns were correctly stressed during silent articulation. We refrained from making predictions for the PP as we suspected that inference and interpretation might be complicated by two factors: (a) by the unusual three-word presentation mode that only allows extracting ERPs time-locked to the onset of the whole phrase (and not to the dative-marked noun alone), and (b) by potential temporal jitter, component overlap and component smearing that might result from inter-item differences in terms of prosodic complexity (3-6 syllables), as well as word classes (determiners vs. adjectives) and resulting conceptual complexity (one vs. two content words).

Verbs were identical across conditions. Critical verbs were monosyllabic and unambiguously marked for agreement with a third person singular subject. In line with the Principle of Rhythmic Alternation, dative case marking in the preceding constituent resulted in distinct prosodic preferences for the verb position. In $Zero_{DAT}$ conditions, an accented monosyllabic noun precedes the verb and an unstressed syllable is preferred at the verb position. In $Schwa_{DAT}$ conditions, a trochaic noun precedes the verb and results in a preference for a prominent

syllable. Thus, monosyllabic verbs occurred in either a metrically prominent position ($Schwa_{DAT}$) or in a non-prominent one ($Zero_{DAT}$). As the polysyllabic word forms of these verbs (e.g., infinitives) are initially stressed, we hypothesized that lexical access would be facilitated if the monosyllabic forms occurred in a position expected to be prosodically prominent, i.e., in $Schwa_{DAT}$ trials.

Subject NPs consisted of a determiner and a noun; determiners were marked for nominative case and indicated subjecthood. The event participant roles assigned by the verbs (e.g., glance, weep, sit, laugh) typically required a human/animate subject, i.e., an animate third-person singular subject NP was expected on the basis of the available verb information. The semantic congruency ($Congruent_{SEM}$ vs. $Incongruent_{SEM}$) of the NP was manipulated by varying the semantic fit between verbs and subjects: semantically incongruent sentences were constructed by replacing the nouns of the subject NP in the $Congruent_{SEM}$ sentences while keeping determiners (i.e., morphosyntactic features) identical. The referents of the $Incongruent_{SEM}$ nouns were less likely actors for the actions described (e.g., "*a robber/priest threatens with a dagger*") and/or violated the animacy restrictions of the verb (e.g., "*a girl/vice sleeps in a bed*"). Inanimate nouns typically expressed abstract concepts (e.g., *greed*, *innocence*, *vice*), lending themselves more readily to a figurative reading than concrete objects (e.g., "*the screwdriver snores*"). The two noun sets were matched for lexical frequency[3] and word length (both paired *t*s < 1; both *p*s > .46), but differed in terms of cloze probability (as assessed in an offline sentence completion task; n=40): congruent nouns were more predictable than incongruous ones (paired $t(47)=3.84$, $p < .001$). Note, however, that the predictability of the critical nouns was very low in both the incongruent condition (range: .00–.05; median: .00) and the congruent condition (range: .00–.63; median: .03). In fact, more than two thirds of the congruent items had a cloze value of .05 or lower, which is typically considered very low contextual constraint. By comparison, a recent study that contrasted the processing of high and medium cloze probability words (Cer-

[3]Lexical frequency was operationalized as the number of occurrences in *Wortschatz*, an online corpus of contemporary written German provided and maintained by the University of Leipzig, Germany (http://wortschatz.uni-leipzig.de).

molacce, Scanella, Faugère, Vion-Dury & Besson, 2014) used item sets with mean cloze values of .95 and .46, respectively. In this respect, the current study differed from many previous ones: as predictability typically requires semantic congruency, the two variables are hard to dissociate experimentally and often confounded when effects of semantic incongruency are investigated.

1.2 Experiment 1a: Sentence judgments of meaningfulness

The critical materials were pretested in a pen-and-paper rating study. Our aims were (1) to ascertain that our sentence variants indeed differed in terms of semantic congruency, and to test for potential effects of (2a) sentence prosody (= Dative) and (2b) genre awareness (= Instruction) on offline evaluation of meaningfulness. We tested the full set of critical stimuli. Note, however, that the sentences did not contain the postcritical adverbial we inserted for the EEG study based on methodological considerations. This ensured that observed differences in meaningfulness were most likely due to the semantic fit between verb and subject. Items were distributed over four lists so that each participant was presented with 12 sentences per condition and saw only one variant from each item set; four constrained randomizations were prepared for each of these lists. In addition to the resulting 48 critical sentences, each list contained 10 control items (5 highly implausible, 5 perfectly plausible) intended to represent the extreme ends of the scale, and enabling us to assess whether participants had understood the task.

Concerning aim (1), we expected a clear effect of semantic congruency with lower average meaningfulness ratings for $\text{Incongruent}_{\text{SEM}}$ vs. $\text{Congruent}_{\text{SEM}}$ sentences. Regarding aim (2a), we expected an effect of Dative on semantic evaluation only if phonological processing fluency is misattributed to perceived meaningfulness (McGlone & Tofighbakhsh, 1999). Our predictions regarding aim (2b) are based on a previous report of increased meaningfulness ratings for poetic vs. randomly generated metaphors (Gibbs et al., 1991). If these earlier findings reflect a categorical intentionality effect, i.e., an effect of intentional composition vs. random combination, we should not find genre-dependent dif-

ferences. If, on the other hand, the search for additional significance forms part of the poetry schema, i.e., if assumed intentionality is gradual, we should likewise find increased meaningfulness ratings in the poetry instruction group, possibly restricted to semantically incongruent sentences prompting genre-specific interpretive strategies (Culler, 1975).

1.2.1 Methods

Participants and procedure

A total of 128 members of the University of Frankfurt community were recruited on campus, all of them adult native speakers of German according to self-report (acquisition started no later than age 3). They were given a campus cafeteria coffee voucher as compensation for participation in the study, which took about 10-15 minutes. No further personal information was collected. Participants received one of two instructions, asking them to rate either "sentences" ($\text{Neutral}_{\text{INST}}$) or "verses" ($\text{Poetry}_{\text{INST}}$) on a 6-point rating scale ranging from 1 = *not meaningful at all*[4] to 6 = *very meaningful*. English translations of the instruction texts are provided in Table 1.3.

Data analysis

We tested for main effects and interactions of Instruction, Dative, and Semantic congruency on ratings using linear mixed effects regression with crossed random effects for participants and items (Baayen, Davidson & Bates, 2008). Reported *p*-value estimates are based on Satterthwaite's approximation for degrees of freedom; *p*-values of multiple comparisons are fdr-corrected. Analysis was carried out in *R* (R Core Team, 2016) using the packages *lme4* (Bates, Mächler, Bolker & Walker, 2015) and *lmerTest* (Kuznetsova, Brockhoff & Christensen, 2017).

[4]We used the term '*sinnvoll*', which is not fully equivalent to English '*meaningful*': the German term conceptually covers both 'to make sense' and 'full of significance or purpose'.

Table 1.3: English translations of the instruction texts for the Neutral and the Poetry group in the rating study. Boldface highlights differences between instructions.

Neutral instruction
Thank you very much for your participation.
This study is part of a research project on **language**.
In this questionnaire you are going to read a number of single **sentences**.
Each of these expresses a situation or state of affairs.
Please read these **sentences** attentively and rate how meaningful they are using the scale next to the **sentence**. The scale ranges from "not meaningful at all" to "very meaningful".
Simply mark the box that, in your opinion, best locates the **sentence** on the scale.
Please make sure to rate all **sentences**.

Poetry instruction
Thank you very much for your participation.
This study is part of a research project on **poetic language**.
In this questionnaire you are going to read a number of single **verses**[5] selected from poems.
Each of these expresses a situation or state of affairs.
Please read these **verses** attentively and rate how meaningful they are using the scale next to the **verse**. The scale ranges from "not meaningful at all" to "very meaningful".
Simply mark the box that, in your opinion, best locates the **verse** on the scale.
Please make sure to rate all of the **verses**.

1.2.2 Results and interim discussion

Results of the meaningfulness rating are summarized in Table 1.4. Analysis confirmed that our critical word substitutions successfully manipulated semantic congruency. All of our Incongruent$_{SEM}$ sentence versions were evaluated as less meaningful than their Congruent$_{SEM}$ counterparts, with a mean difference of 1.94 (SD = 0.91) on the 6-point scale [t(5968.9) = -59.92, $p < .001$]. Analysis further revealed that participants assigned more meaning to the sentences if they believed to be reading poetic verses [t(127.5) = 3.08, p = .003].

These simple effects were dissimilar within factor levels, as indicated by the significant interaction of Instruction and Semantic congruency [F(1,5968.9) = 8.35, p = .004]. Resolving the interaction by Semantic congruency, we found only a trend towards a genre-induced increase in the perceived meaningfulness of Congruent$_{SEM}$ sentences [t(160.3) = 1.96, p = .052]; ratings for Incongruent$_{SEM}$

sentences, by contrast, were reliably increased by the poetry instruction [t(160.3) = 3.86, $p < .001$], see Figure 1.1. Meaningfulness ratings were insensitive to the Dative manipulation [F(1,123.5) < 1].

Table 1.4: Results of the meaningfulness rating. Condition means (+*SD*s); 1 = not meaningful at all, 6 = very meaningful

	$Zero_{DAT}$		$Schwa_{DAT}$	
	$Congruent_{SEM}$	$Incongruent_{SEM}$	$Congruent_{SEM}$	$Incongruent_{SEM}$
$Neutral_{INST}$	5.25 (0.65)	3.20 (0.74)	5.23 (0.73)	3.21 (0.81)
$Poetry_{INST}$	5.41 (0.51)	3.62 (0.70)	5.46 (0.49)	3.54 (0.75)

Supplementing the findings of Gibbs and colleagues (1991), these results suggest that the poetry schema adds to the categorical intentionality effect: reading poetry entails the controlled search for additional significance beyond plain sense (e.g., Culler, 1975; Thorne, 1988), particularly if simple composition fails to yield a plausible/coherent propositional representation of sentence meaning.

1.3 Experiment 1b: ERP responses during sentence reading

We expected to find modulations of the N400, an ERP component that has traditionally been linked to semantic processing (see Kutas & Federmeier, 2011 for a review). Based on previous ERP results on the processing of semantic incongruities (e.g., Federmeier & Kutas, 1999; Kutas & Hillyard, 1980), and based on the results of Experiment 1a, we predicted increased N400 responses to $Incongruent_{SEM}$ vs. $Congruent_{SEM}$ subject noun phrases.

Crucially, amplitude and topography of the N400 – and to a lesser degree also its temporal characteristics – are sensitive to phonological processes that facilitate (or delay) lexical access, such as onset priming, rhyme priming, and prosodic priming. This has been demonstrated in auditory word, sentence, and text processing (e.g., Bohn et al., 2013; Connolly & Phillips, 1994; Henrich et al., 2014; Obermeier et al., 2015; Praamstra, Meyer & Levelt, 1994;

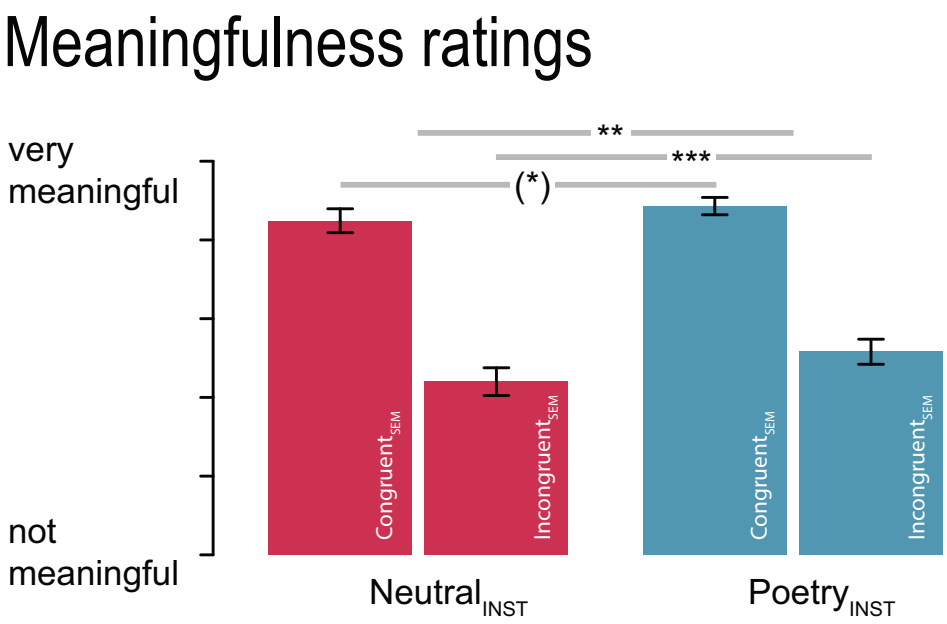

Figure 1.1: Results of Exp. 1a: Average meaningfulness ratings for $\text{Incongruent}_{\text{SEM}}$ vs. $\text{Congruent}_{\text{SEM}}$ sentences by Instruction ($\text{Poetry}_{\text{INST}}$ vs. $\text{Neutral}_{\text{INST}}$); n=64 each). Sentences were rated on a 6-point-scale ranging from "*not meaningful at all*" to "*very meaningful*". Error bars indicate 95% confidence intervals; asterisks indicate statistical significance of planned comparisons: n.s. = not significant; (*) $p < .08$; * $p < .05$; ** $p < .01$; *** $p < .001$.

Praamstra & Stegeman, 1993; Rothermich, Schmidt-Kassow & Kotz, 2012; Rugg, 1984a, 1984b; Rugg & Barrett, 1987). For instance, Rothermich and colleagues reported modulations of N400 amplitudes depending on prosodic context, which they interpreted as evidence that the predictability of stress locations facilitates lexico-semantic processing. Importantly, N400-like prosodic negativities have been reported for silent reading as well (e.g., Luo & Zhou, 2010; Magne, Gordon & Midha, 2010), in line with mounting evidence that implicit prosody (Fodor, 1998) co-determines the comprehension of written text (e.g., Bader, 1998; Breen & Clifton, 2011; Kentner, 2012; Kentner & Vasishth, 2016; Kriukova & Mani, 2016; Steinhauer & Friederici, 2001; Stolterfoht et al., 2007; see Breen, 2014 for a recent review). Therefore, we expected that strict rhythmical alternation ($\text{Schwa}_{\text{DAT}}$) would allow predicting upcoming stress locations and thus facilitate lexico-semantic processing, resulting in reduced

N400 responses (compared to $Zero_{DAT}$) at the sentence constituent following the morpho-phonological manipulation, i.e., at the verb.

Importantly, the N400 is also sensitive to task demands and varies as a function of the depth of processing (e.g. Bentin, Kutas & Hillyard, 1993; Chwilla, Brown & Hagoort, 1995; Rugg, 1984a; West & Holcomb, 2000). Connecting this observation to Hanauer's (1998b) hypothesis that the poetry reading mode modulates the level of processing, we further expected to observe distinct electrophysiological responses to morpho-phonological and semantic features in the two groups, if readers' poetry schema entails adapted processing routines for either of these domains. More specifically, if poetry readers attend to sound recurrences in the assumption that these are not random, the poetry condition should exhibit more efficient use of prosodic cues and greater facilitation of rhythmic and/or lexico-semantic processing in favorable prosodic contexts (interaction effect: Instruction x Dative). This effect should be accompanied by a general reduction of N400 responses at both the verb and the subject NP (main effect: Instruction), if the genre-induced increase in attention to sound reduces the depth of semantic processing, parallel to the reduction of semantic effects due to task requirements (Bentin et al., 1993; Chwilla et al., 1995; Rothermich et al., 2012). If—as suggested by the results of the rating experiment—unusual conceptual combinations are more expectable in poetic contexts, the comprehension system should downgrade contextual constraint on semantic access and lead to a reduced semantic N400 effect at the NP (interaction effect: Instruction x Semantic congruency). Alternatively, the interaction effect observed in offline ratings might be reflected in ERP differences in a late positive component, i.e., the P600. This component was long believed to be a marker of syntactic processing cost (e.g., Osterhout & Holcomb, 1992) but it also occurs in other contexts that are potentially relevant to the present study. First, the P600 is assumed to index the task-dependent evaluation of well-formedness (e.g., Bohn et al., 2013; Bornkessel-Schlesewsky & Schlesewsky, 2008). Second, P600 effects have been observed as part of a biphasic N400/P600 pattern in response to semantic incongruency and to animacy violations (e.g., Kutas & Hillyard, 1980; Szewczyk & Schriefers, 2011). Finally, the P600 is assumed

to reflect a late pragmatic processing stage that follows semantic processing proper (as reflected in the N400), for instance in inferencing (Burkhardt, 2007), or in the computation of metaphorical meaning (e.g., Coulson & Van Petten, 2002; Weiland, Bambini & Schumacher, 2014). More generally, the P600 and late positivities are believed to reflect controlled stimulus evaluation and the resolution of conflicts from prior processing streams (e.g., Bornkessel-Schlesewsky & Schlesewsky, 2008).

Our predictions concerning the probe task were as follows: if the processing system adapts attention allocation and memory encoding to the input category, i.e, the genre, then a memory effect should already be detectable in immediate recognition, leading to enhanced probe task performance for the poetry group ($Poetry_{INST} > Neutral_{INST}$); failure to find differences between the groups in terms of probe task performance ($Poetry_{INST} = Neutral_{INST}$) is compatible with the alternative explanations that the memory effects observed in the delayed tasks of previous studies reflect slower reading and thus longer encoding (Hanauer, 1998a), or that they hinge on enhanced retrieval due to facilitating sound recurrences (Tillmann & Dowling, 2007). Our expectations concerning the Dative conditions depended on whether performance impairment due to infrequent morphology outweighs the benefits of facilitated prosodic processing ($Zero_{DAT} > Schwa_{DAT}$) or vice versa ($Schwa_{DAT} > Zero_{DAT}$). Furthermore, if the poetry schema licenses morphological deviations, we should find reduced impairment or enhanced facilitation of behavioral performance in this group (interaction: Instruction x Dative).

1.3.1 Methods

Participants

Forty-eight monolingually raised native speakers of German (acquisition started no later than age 3) were recruited from the University of Mainz community and were paid 7 EUR per hour for participation in the experiment. All participants reported normal or corrected-to-normal vision and no known neurological or reading disorders, and all were right-handed as assessed with an abridged Ger-

man version of the Edinburgh handedness inventory (Oldfield, 1971). Prior to the experiment, participants were informed about the experimental procedures and gave written consent. The data from four additional participants had to be excluded due to excessive body and/or eye movement artifacts. Participants were pseudo-randomly assigned to one of two instruction groups ($\text{Neutral}_{\text{INST}}$ or $\text{Poetry}_{\text{INST}}$). Care was taken that the groups matched in terms of the number of participants (n = 24 each) as well as sex (12 females per group) and age [$\text{Neutral}_{\text{INST}}$: M = 23.63, SD = 2.79; $\text{Poetry}_{\text{INST}}$: M = 24.83, SD = 3.57; two-sample t-test: $t(46) < 1$].

Procedure

The 192 critical sentences resulting from the 48 item sets were distributed over two lists, each containing 96 critical sentences (2 chosen from each item set to total 24 sentences per condition) and 48 filler sentences[6]. Sentences that were taken from the same item set included one $\text{Congruent}_{\text{SEM}}$ and one $\text{Incongruent}_{\text{SEM}}$ continuation, with combinations of the two Dative variants rotated across items. Care was taken that at least 25 trials intervened between these pairs, and their order of occurrence was balanced across the two constrained randomizations we prepared for each list. All sentences conformed to the constituent structure described in section 1.1.1.

Participants received written instructions on a sheet of paper prior to the experiment; the experimenter ensured that instructions had been read and understood. Translations of the instruction texts are provided in Table 1.5. All experimental procedures were ethically approved by the Ethics Council of the Max Planck Society, and were undertaken with written informed consent of each participant.

[6]Fillers were identical to the critical materials in terms of syntactic structure. They also contained implausible (50%) and metrically regular (50%) sentences. Contrary to the critical materials, fillers showed a greater variety of semantic and metrical variation. For instance, in filler sentences, implausibility sometimes arose PP-internally from semantic mismatch between adjective and noun (instead of semantic fit between verb and subject). Metrical variation included ternary meters (e.g., dactyls) and polysyllabic verbs with different word stress patterns.

Table 1.5: English translations of the instruction texts for the Neutral and the Poetry group in the ERP study. Boldface highlights differences between instructions.

Neutral instruction
Thank you very much for your participation in our **language** study.
You are going to read a number of single **sentences** on the screen.
Please read these **sentences** attentively.
Before each **sentence** a little star (*) indicates that the first part of the **sentence** is going to appear.
After reading a **sentence**, you are going to be presented with a word.
Your task is to decide whether this word had occurred in the **sentence** or not.

Poetry instruction
Thank you very much for your participation in our **poetry** study.
You are going to read a number of single **verses selected from poems** on the screen.
Please read these **verses** attentively.
Before each **verse**, a little star (*) indicates that the first part of the **verse** is going to appear.
After reading a **verse** you are going to be presented with a word.
Your task is to decide whether this word had occurred in the **verse** or not.

Experimental sessions were conducted in a dimly lit and sound-attenuated room. Participants were seated comfortably at a distance of about 100 cm from a 17" computer screen. Following a short practice session, the experiment consisted of four blocks of 36 sentences/verses each (approx. 8 minutes per block). Participants were encouraged to take short breaks (1-2 minutes) between blocks to relax and rest their eyes, as we had asked them to avoid movements and eye blinks during sentence presentation.

Sentences were presented constituent-wise in a 29-point font in the center of the display. Each trial began with the presentation of a fixation asterisk for 1000 ms, followed by a 200 ms inter-stimulus interval (ISI). Presentation rates were adapted to constituent size (e.g., Bornkessel, Fiebach & Friederici, 2004; Schumacher & Hung, 2012): 450 ms for the PP (three words), 300 ms for the verb (one word) and 400 ms for the NP (two words), each followed by a 200 ms ISI. Presentation durations for the post-critical constituents also varied with length: one word = 300 ms (+200 ms ISI), two words = 400 ms (+200 ms ISI), three words = 450 ms (+200 ms ISI). Each trial ended with 500 ms of blank screen, after which participants performed a probe recognition task (cued by a

question mark). In the probe task, participants were presented with a word, and indicated whether it had occurred in the sentence or not by pressing one of two pushbuttons on a hand-held gamepad; participants did not receive feedback on their performance during the experiment. The number of expected yes/no responses[7] was balanced for each participant, as was the number of probes targeting each of the constituents. The assignment of left and right buttons to yes/no responses was counterbalanced across participants and within groups. The trial ended once a valid button press had been registered, or when the maximal response time (set at 5000 ms) had elapsed. After a further 1500 ms of blank screen the next trial started. A schematic depiction of this trial structure is provided in Table 1.6.

Table 1.6: Schematic trial structure (*bs* = blank screen) and presentation durations in msec.

Pre-stimulus		Sentence presentation								Probe task			
*	*bs*	**PP**	*bs*	**verb**	*bs*	**NP**	*bs*	**ADV**	*bs*	**?**	*bs*	**probe**	*bs*
1000	200	450	200	300	200	400	200	300-450	500	300	200	<5000	1500

EEG recording and preprocessing

The EEG was recorded from 26 Ag/AgCl electrodes (ground: AFZ) fixed at the scalp by means of an elastic cap (Easycap GmbH, Herrsching, Germany). Four additional electrodes monitored the electro-oculogram (EOG) at the outer canthus of each eye (horizontal EOG) and above and below the participant's right eye (vertical EOG). EEG and EOG channels were amplified by means of a BrainAmp amplifier (Brain Products, Gilching, Germany) and digitized with a sampling rate of 500 Hz. Recordings were referenced to the right mastoid but re-referenced to linked mastoids offline. Electrode impedance was kept below 5 kΩ during the experiment. We applied a 0.3–20 Hz band-pass filter to the raw EEG data offline to eliminate slow signal drifts and high frequency noise; the filter

[7] No-response probes were either phonologically (50%) or semantically (50%) related to the constituents targeted by the corresponding yes-response probes.

avoids potential stimulus-independent differences between conditions without performing a baseline correction which potentially carries over transient baseline differences into the critical region (for a motivation of this approach, cf. Schumacher & Hung, 2012; Wolff, Schlesewsky, Hirotani & Bornkessel-Schlesewsky, 2008; see also Widmann, Schröger & Maess, 2015). We extracted ERP epochs from 200 ms before constituent onset to 1000 ms post onset before calculating single-trial averages for the latency bins of interest. Incorrectly answered and timed-out (> 5000 ms) trials, as well as trials containing EEG or EOG artifacts (EOG rejection criterion: 40 µV; rejection rates: verb 1.2%, NP 1.9%), were excluded from further analysis (about 6% of the original data).

Data analysis

For the statistical analyses of behavioral and ERP data, we tested for fixed main and interaction effects of Instruction, Dative and Semantic congruency using linear mixed-effects regression. For the ERP data of the verb position, where semantic information was not yet available, we only tested for fixed main and interaction effects of Instruction and Dative. Models contained crossed random effects for participants and items (Baayen et al., 2008). Parsimonious random effect structures were determined using the forward-fitting algorithm implemented in the *R* package *LMERConvenienceFunctions* (Tremblay & Ransijn, 2015); the alpha level for the log likelihood ratio tests was set at 0.1. In order to facilitate interpretation and comparison, we report *F*- and *p*-value estimates based on the Satterthwaite approximation for degrees of freedom. Interactions were resolved hierarchically, controlling the false discovery rate at each level of analysis (Benjamini & Hochberg, 1995). All analyses were carried out in *R* (R Core Team, 2016) using the packages *lme4* (Bates, Mächler, Bolker & Walker, 2015), *LMERConvenienceFunctions* (Tremblay & Ransijn, 2015), *lmerTest* (Kuznetsova et al., 2017) and *multcomp* (Hothorn, Bretz and Westfall, 2008).

Behavioral data analysis

Reaction times: We removed outliers exceeding 2.5 *SD*s from both the partici-

pant's and the item's mean before analyzing log-transformed reaction times of correctly answered trials. On the basis of these criteria, 7.6% of the original data points were excluded. The subsample was biased in the sense that, due to response accuracy, more $\text{Schwa}_{\text{DAT}}$ trials were excluded than Zero_{DAT} trials (Chi-squared test for given probabilities: $\chi^2(1) = 16.02$, $p < .001$).
Response accuracy: We used logistic regression to test for main and interaction effects of Instruction, Dative and Semantic congruency on response accuracy (correct/incorrect) .

ERP data analysis
The ERP data analysis was carried out by fitting separate models for midline (FZ, FCZ, CZ, CPZ, PZ) and lateral electrodes (F3, F7, FC1, FC5, CP1, CP5, P3, P7, F4, F8, FC2, FC6, CP2, CP6, P4, P8). Outlying values exceeding +/- 20 microvolts were excluded from the analysis (approx. 1% data loss). For the lateral electrodes, four regions of interest were defined via combinations of the spatial factors Hemisphere (left/right) and Region (anterior/posterior). Contrast-coded spatial factors were included in the fixed effects term of the respective models, allowing for all interactions. Note that we do not report main effects and/or two-way interactions of spatial factors.

Time window selection: We predicted ERP differences in the N400 latency range (300–600 ms). The actual time windows for the analyses were chosen from this pre-determined latency range upon visual inspection of the grand average plots.

1.3.2 Results

Behavioral results

The behavioral data (see Table 1.7) indicate that participants generally processed the sentences/ verses attentively. Mean accuracy was 96.05%; mean response latency was 966 ms (1042 ms before outlier removal).

Table 1.7: Mean reaction times (RT) in msec and error rates per condition; standard deviations (*SD* are given in parentheses)

		$Zero_{DAT}$		$Schwa_{DAT}$	
		$Congruent_{SEM}$	$Incongruent_{SEM}$	$Congruent_{SEM}$	$Incongruent_{SEM}$
$Neutral_{INST}$	*RT*	1003 (192)	976 (176)	971 (174)	1017 (194)
(*n* = 24)	*error rate*	1.57 (2.41)	2.78 (2.92)	5.90 (5.20)	5.22 (5.26)
$Poetry_{INST}$	*RT*	953 (224)	937 (192)	945 (202)	961 (197)
(*n* = 24)	*error rate*	2.78 (3.82)	3.32 (3.26)	3.82 (3.23)	4.69 (5.11)

Reaction Times

Reaction times were on average 44 ms shorter for the $Poetry_{INST}$ group (M = 949 ms, SEM = 36 ms) than for the $Neutral_{INST}$ group (M = 993 ms, SEM = 41 ms), though this difference was not reliable [$F(1,47.9) < 1$]. Analysis further revealed that the effect of semantic congruency differed between levels of Dative [Dative x Semantics: $F(1,4168.2) = 11.72$, $p < .001$]. If sentences contained archaic $Schwa_{DAT}$ case marking, reaction times were higher in $Incongruent_{SEM}$ vs. $Congruent_{SEM}$ trials [$t(4168) = 3.00$, $p = .005$]. $Zero_{DAT}$ trials showed a trend towards a reaction time advantage for $Incongruent_{SEM}$ vs. $Congruent_{SEM}$ trials [$t(4167) = -1.83$, $p = .067$]; see Figure 1.2, left panel. Neither the main effect of Dative nor the interaction effect of Instruction and Dative approached significance (all Fs < 1, all ps > 0.4).

Accuracy

There was a main effect of Dative on response accuracy: error rates were higher in the $Schwa_{DAT}$ conditions than in the $Zero_{DAT}$ conditions [$z = –3.62$, $p < .001$], but this effect was dissimilar across levels of Instruction, as indicated by a significant interaction [$z = 2.06$, $p = .039$]. Planned comparisons revealed that archaic $Schwa_{DAT}$ nouns increased error rates in the $Neutral_{INST}$ condition [$z = –3.62$, $p < .001$] but not in the $Poetry_{INST}$ condition [$z = –0.99$, $p = .324$]; see Figure 1.2, right panel.

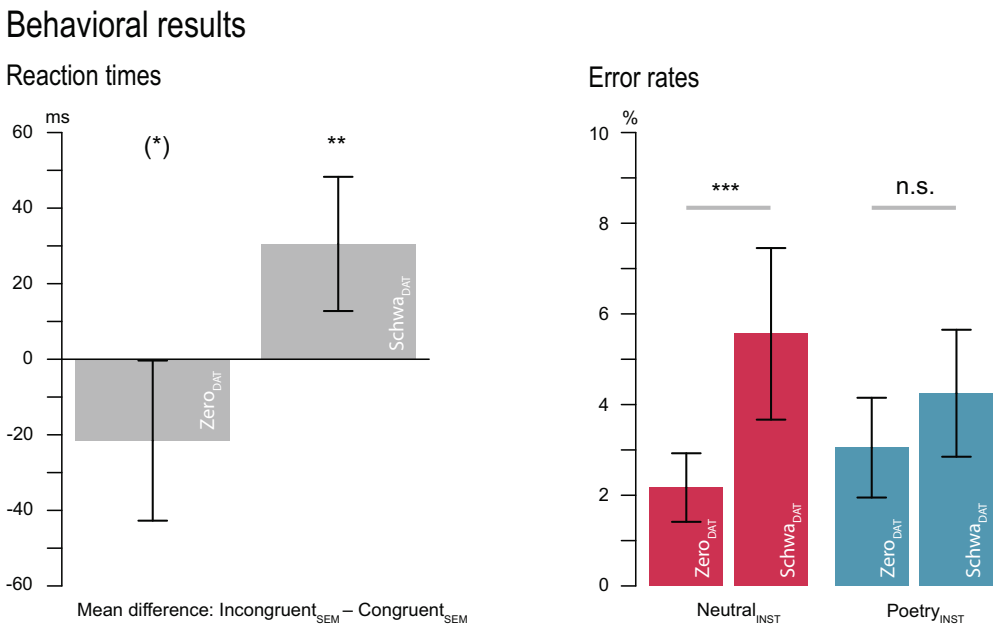

Figure 1.2: Exp. 1b: Behavioral results: Mean reaction time differences ($Incongruent_{SEM}$ – $Congruent_{SEM}$) in ms for the dative conditions (left panel), and mean error rates for the dative conditions by instruction (right panel). Error bars indicate 95% confidence intervals; asterisks indicate statistical significance of planned comparisons: n.s. = not significant; (*) $p < .08$; * $p < .05$; ** $p < .01$; *** $p < .001$.

ERP results

Verb

Time window 350–600 ms: Verbs following $Zero_{DAT}$-PPs elicited an increased negativity relative to verbs following $Schwa_{DAT}$-PPs; the effect was widely distributed and most pronounced over left and central regions (see Figure 1.3). Statistical analysis confirmed a main effect of Dative at midline electrodes [$F(1,45.971) = 12.22$, $p = .001$, $t(45.9) = -3.97$] and lateral sites [$F(1,46) = 15.77$, $p < .001$, $t(45.9) = -3.97$]. The lateral effect was qualified by an interaction with Hemisphere [$F(1,69917) = 7.88$, $p = .005$] reflecting that the dative effect was more pronounced over the left [$t(59.2) = -4.70$, $p < .001$] than over the right hemisphere [$t(59.2) = -2.76$, $p = .008$]. Contrary to our expectations, none of the effects differed between instruction groups [all $ps > .25$]. The repetition of the

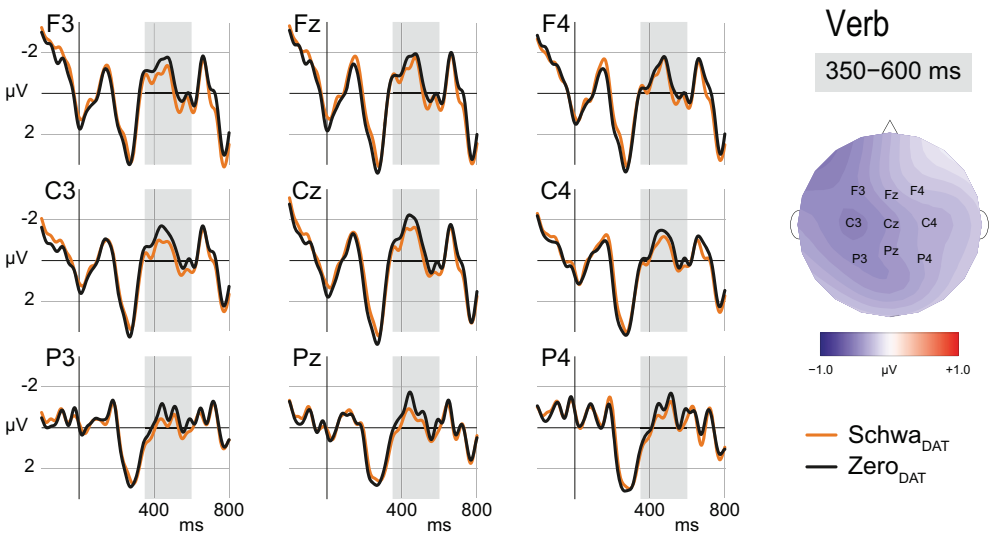

Figure 1.3: Exp. 1b: Grand-average ERP responses to the verb. Canonical $Zero_{DAT}$ conditions show an increased negativity (350-600ms) compared to non-canonical $Schwa_{DAT}$ conditions.

critical verbs did not modulate these effects of the morpho-phonological manipulation [all *p*s > .36].

Subject NP

Congruency conditions diverged around 400 ms after NP onset and showed a bilateral posterior negativity (N400) for incongruent vs. congruent NPs (see Figure 1.4). This effect was followed by late positive components (P600) whose onset latency differed between Instruction conditions (see Figure 1.5). In the neutral condition, incongruent NPs elicited a P600 between 600–700 ms after NP onset that was most pronounced over central and frontal sites. The poetry condition, in contrast, showed a sustained posterior negativity with an offset latency of 700 ms, followed by a centro-parietal positivity (P600) between 750–900 ms.

Time window 400–600 ms: There was no semantic effect at midline electrodes [$F(1,76.5) < 1$, $p = .440$], but we observed an interaction of Semantic

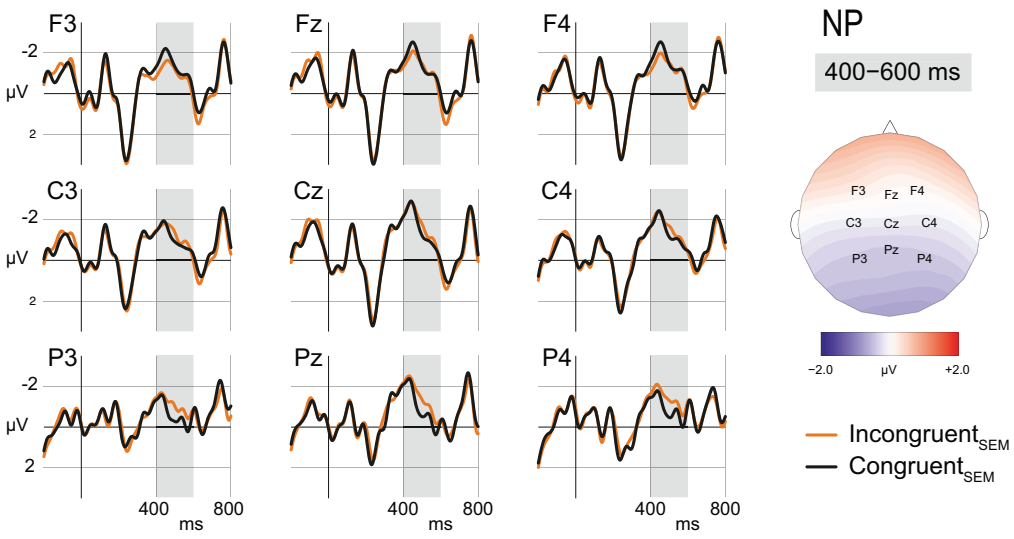

Figure 1.4: Exp. 1b: Grand-average ERPs responses to the subject NP. Incongruent$_{SEM}$ conditions show an increased bilateral posterior negativity (400-600ms) compared to non-canonical Congruent$_{SEM}$ conditions.

congruency and Region for lateral sites [$F(1,68982) = 82.90$, $p < .001$]. Resolving the interaction by Region verified a semantic negativity at posterior [$t(48.1) = –3.44$, $p = .002$], but no effect at anterior sites [$t(48) = 1.60$, $p = .145$].

Time window 600–700 ms: At midline electrodes, there was an interaction effect of Instruction and Semantic congruency [$F(1,21523.3) = 24.12$, $p < .001$]. Whereas the neutral condition showed more positive-going ERPs in response to incongruent vs. congruent NPs [$t(73) = 3.37$, $p = .002$], the poetry condition showed no difference [$t(71.5) = -0.90$, $p = .371$]. Laterally, this effect differed between Regions [$F(1,68911.3) = 4.73$, $p = .030$]. Absent in the posterior region [$F(1,48) = 2.15$, $p = .149$], an interaction effect of Instruction and Semantic congruency emerged over the anterior region [$F(1,34360) = 28.85$, $p < .001$], again reflecting more positive-going ERPs for incongruent vs. congruent NPs in the neutral condition [$t(64.7) = 3.30$, $p = .003$], and no difference in the poetry condition [$t(63.6) = 0.30$, $p = .762$].

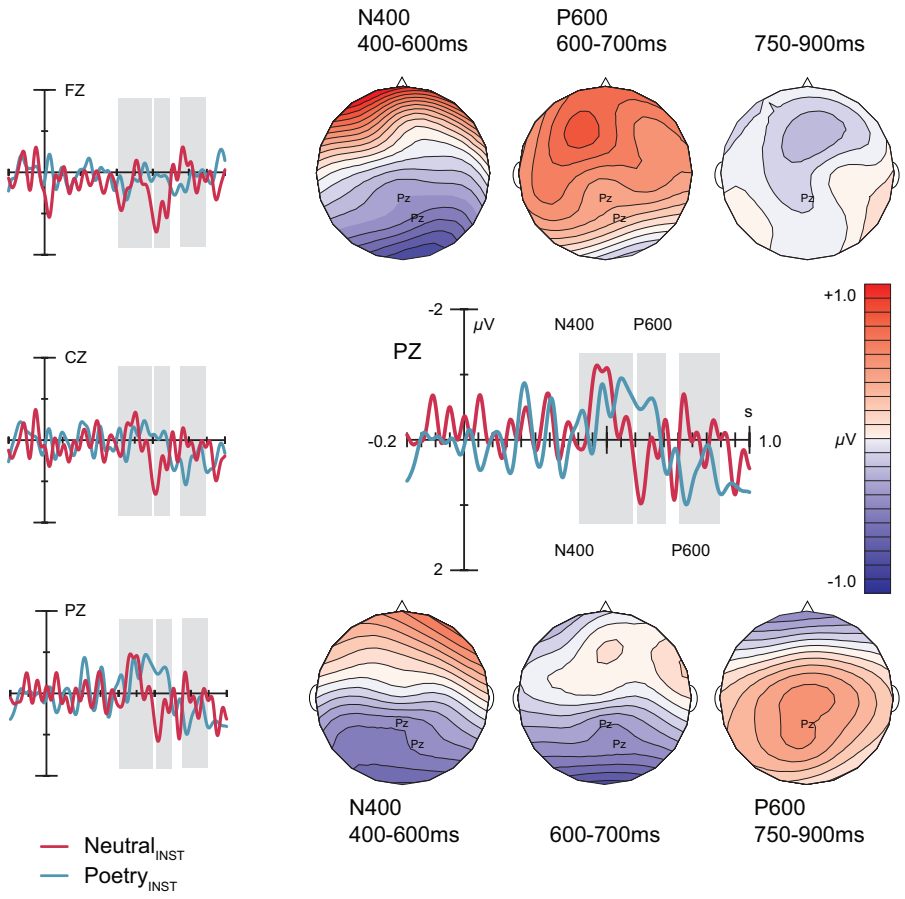

Figure 1.5: Exp. 1b: Instruction-wise comparison of semantic ERP effects ($Incongruent_{SEM}$–$Congruent_{SEM}$) at the subject NP. Difference waves (left column) and difference maps illustrate semantic congruency effects for the $Neutral_{INST}$ (black lines; upper row of maps) and for the $Poetry_{INST}$ (red lines; lower row of maps) conditions. Both conditions show a posterior negativity (N400) between 400-600ms. The late positive response (P600) occurs earlier in the $Neutral_{INST}$ condition (600-700ms) than in the $Poetry_{INST}$ condition (750-900ms).

Time window 750–900 ms: At midline sites, there was an interaction effect of Instruction and Semantic congruence [$F(1, 21433.4) = 15.79$, $p < .001$]. The neutral condition showed no semantic effect in this latency range [$t(75.3) = -0.97$, $p = .335$], but in the poetry condition ERPs were more positive-going in response to incongruent vs. congruent NPs [$t(73.6) = 2.58$, $p = .024$]. At lateral

electrodes, we found an interaction effect of Instruction and Semantic congruency [$F(1,69233.4) = 17.33$, $p < .001$]. Resolution revealed a trend towards more-positive going ERPs in response to incongruent vs. congruent NPs in the poetry group [$t(58.8) = 2.10$, $p = .080$], and no difference between semantic conditions in the neutral group [$t(59.5) = –0.53$, $p = .595$]. Resolution of a significant interaction of Semantic congruency and Region [$F(1,69147) = 8.87$, $p = .002$] revealed no reliable effects [all uncorrected ps > .10].

1.4 Discussion and conclusion

Focusing on poetry and poetry-typical linguistic features, the present study aimed to investigate the effects of genre awareness on the processing and evaluation of isolated sentences. Differential instruction of participant groups allowed us to dissociate top-down-controlled genre effects from input-driven automatic processes.

In an offline rating study (Exp. 1a), we validated our semantic manipulation and tested for effects of genre awareness on the perceived meaningfulness of the stimuli. Poetry categorization resulted in increased ratings of meaningfulness for semantically incongruent sentences.

In an ERP study (Exp. 1b), we tested whether similar effects are detectable already in online processing. We expected (1) a prosodic N400-like effect at the verb due to facilitated lexico-semantic processing resulting from prosodic expectations built up during the processing of the sentence-initial constituent, and (2) a semantic N400 effect at the subject NP indexing the processing of incongruent meaning. We further hypothesized that (3a) the prosodic effect would be reinforced and that (3b) the semantic one would be attenuated by the poetry instruction. The results supported hypotheses (1) and (2) while hypotheses (3a) and (3b) were disconfirmed by the ERP data. The expected interaction of genre awareness and semantic congruency (3b) was reflected in the latency of the P600 response to incongruent vs. congruent NPs rather than in top-down modulations of N400 amplitudes. The interaction of genre awareness

and morpho-phonological alternation (3a) was found in behavioral measures: rather than increasing facilitation, poetry instruction neutralized the adverse effect of stylistically marked case suffixes on response accuracy.

The probe task tested whether poetry categorization affects surface structure memory immediately after sentence wrap-up. Such an effect would provide strong evidence for the hypothesis that genre awareness modulates the level of processing and thus the resulting memory traces, which, in turn, could (partially) account for the enhanced surface structure recall and recognition reported for poetry versus other text types. Results did not show the predicted general advantage for the poetry group and thus failed to support the level-of-processing account, at least in its strongest, aprioristic version. Regardless of the instruction, response latencies were delayed after processing sentences with combined morphological and semantic deviations. Response accuracy differed between instruction groups, reflecting that error rates increased with the presence of slightly archaic case markers when no genre had been specified, but not when sentences had been identified as lines of poetry. Most likely, this effect does not result from a modulation of the level of processing (which predicts a main effect of genre awareness), but merely reflects that stylistically marked case suffixes are more frequent and expectable in poetry than in most other registers. The results thus suggest that previously reported genre effects on surface structure memory are due to textual variables that were absent in the current study. First, the graphic layout of poetic texts seems to enhance the retrieval of surface structure representations after reading poetry, possibly by providing spatial retrieval cues (Hanauer, 1998b). Second, the high degree of interlinear phonological patterning in poetic texts, e.g., line-final rhymes and recurrent meter, seems to increase memorability, as suggested by Tillman and Dowling (2007). In this latter view, systematic sound patterns serve as effective retrieval cues in actual performance (Rubin, 1995), but also in experimental recall and recognition paradigms (Hanauer, 1998a; Lea, Rapp, Elfenbein, Mitchel & Romine, 2008; Rubin, Ciobanu & Langston, 1997; Tillman & Dowling, 2007; van Peer, 1990). Note that this explanation is compatible with a level-of-processing account un-

der the additional assumption that the depth of processing varies as a function of accumulating genre cues like systematic sound patterning (Zwaan, 1993).

Effects of archaic inflection: Behavioral performance and prosodic N400
As pointed out in the introduction, historical inflections provide poets with creative options for balancing the metrical structure of a line. To contemporary readers, however, they constitute morpho-phonological oddities, as indexed by the decline in subsequent probe task performance when no register/genre information had been provided. It seems that the sentence-initial presence of these infrequent word forms interferes with the behavioral task, possibly by introducing additional uncertainty concerning the exact form of the probe word. Importantly, this penalty diminished for participants in the poetry group. To us, this suggests that the poetry readers were better equipped to deal with poetry-typical, i.e., register-dependent, grammatical deviations while performing the task than were readers without previous text type information. Most likely, the genre instruction cued and activated the grammatical register associated with 18th/19th century lyrical poetry which is arguably the most canonical and prototypical representative of the German poetic tradition.

Notably, the morpho-phonological manipulation that impaired behavioral performance in one of the groups facilitated prosodic processing of the following word. In line with our predictions, weobserved a reduced negativity in response to verbs following non-canonically case-marked NPs, resulting in a left-lateralized N400-like effect. It is unlikely that the observed effect is an instance of a semantic N400 that could be accounted for by semantic unification or contextual fit (e.g., Brown & Hagoort, 1993; Hagoort, Baggio & Willems, 2009). First, the scalp distribution differed from the classic semantic N400 effect (Federmeier & Kutas, 1999). More importantly, our design allowed us to test the effect of the morpho-phonological manipulation on identical words embedded in semantically identical contexts within participants. But could this effect merely be a late positivity in response to the morphological archaism in the preceding constituent? We assume that this is not the case for two reasons. First, the late occurrence of the effect (approx. 1000-1250 ms after PP onset) makes

such an interpretation questionable. Second, the case suffix does not constitute an ungrammaticality, i.e., it is a register-dependent stylistic variant, not a violation. Providing licensing register/genre information should thus modulate an effect that supposedly indexes the ill-formedness of a register-dependent alternation. Recall that the behavioral results seem to confirm that the poetry schema indeed licenses these historical inflections. Therefore, we assume that the observed negativity indeed indexes the relative ease of lexico-prosodic processing, i.e., of accessing and integrating a lexeme whose word-prosodic template matches rhythm-based predictions (Henrich et al., 2014; Rothermich et al., 2012). This result aligns with earlier reports of phonological N400 effects in studies using auditorily and visually presented word pairs (Praamstra & Stegeman, 1993; Praamstra et al., 1994; Rugg, 1984a, 1984b; Rugg & Barrett, 1987), single sentences (e.g., Luo & Zhou, 2010; Rothermich et al., 2012), and lyrical stanzas (Obermeier et al., 2015). If this interpretation is correct, our data extend the sentence-level effect of implicit prosody to comparatively subtle differences in morpho-phonological context, showing that the historical inflections frequently found in poetry successfully induce such differences.

Crucially, our findings do not support the idea that genre expectations affect prosodic processing in reading immediately and independent of textual variables. Previously reported effects of meter on poetry processing thus most likely reflect input-driven predictions based on phonological recurrence across lines. This interpretation is still compatible with the view that poetry reading affects top-down control of attention to phonological form. For the poetry group, our out-of-context lines thus correspond to the first line of a poem, where no metrical pattern has been established yet. While readers might expect a metrical pattern to be present, it is unclear, at this stage, which one it will turn out to be. Readers most likely extract the systematic sound patterning instantiated in a text incrementally when reading or hearing the first few lines or stanzas of a poem. This interpretation is in line with Smith's (1968:13) proposal that poetic structure is "an inference which we draw from the evidence of a series of [processing] events". In this view, adaptations of prosodic processing are text-specific and recurrence-based rather than genre-specific and aprioristic.

Taken together, these results suggest that the same deviant morphological feature may impair and facilitate distinct stages of stimulus processing. This constitutes initial evidence from comprehension in support of the assumed trade-off between morphological deviation and prosodic regularity in poetic language (Rice, 1997). Our findings further suggest that genre awareness moderates the impairment of comprehension while leaving the facilitation intact.

Effects of semantic (in)congruency

Offline sentence evaluation confirmed the semantic congruency difference between sentence variants, and revealed a reliable interaction of instruction and semantics: genre categorization increased ratings of meaningfulness for incongruent sentences but not for congruent ones. These findings refine and extend the intention-based meaningfulness effect for metaphors (Gibbs et al., 1991) to non-metaphorical semantic incongruities. Rather than an intentional/non-intentional distinction, the present findings most likely reflect the genre-dependent aspect of the controlled search for hidden significance, and thus support a gradient conception of intentionality. These findings are in line with a two-step conception of genre-dependent semantic interpretation: upon failure to derive a coherent semantic representation by semantic composition (step one), poetry readers search for additional, hidden significance (step two).

Confirmed in ratings of meaningfulness, the semantic congruency difference also affected online sentence processing. In the ERPs, we observed a bilateral posterior N400 in response to semantically incongruent vs. congruent NPs. While the effect showed some characteristics that contrast with the “standard N400 effect”, i.e., a broadly distributed centro-parietal negativity between 250-550ms (Kutas & Federmeier, 2011), its direction, its occurrence in the N400 latency range, as well as its familiar sensitivity to the semantic manipulation demonstrate clearly that we are dealing with an instance of a semantic N400.

As the N400 is highly sensitive to cloze probability (Kutas & Hillyard, 1984), we assume that the late onset of the effect most likely reflects the lack of lexical predictability in both congruency conditions. At the phrasal level, semantic congruency effects on the N400 differ from effects of predictability in terms of

magnitude, latency and topography: while predictability differences elicit the familiar robust centro-parietal negativity, the N400 effects elicited by semantic incongruency are weaker and more variable in terms of scalp distribution (Lau, Namyst, Fogel & Delgado, 2016). Furthermore, the restricted scalp distribution and the reduced magnitude observed here likely also reflect the large proportion of abstract nouns employed in the semantically incongruent conditions (e.g., *greed*, *grief*, *scorn*, *innocence*, *vice*, *curiosity*): abstract nouns are (a) known to elicit less pronounced N400 responses than concrete ones, typically with a more posterior distribution (Kounios & Holcomb, 1994; West & Holcomb, 2000), and (b) judged to be worse event instigators than concrete ones and thus less likely to be realized as the grammatical subject, leading to less pronounced N400 responses if animacy expectations are not met (Frenzel, Schlesewsky & Bornkessel-Schlesewsky, 2015). Finally, the sentence contexts in the present study did not exploit German native speakers' high sensitivity to animacy in sentence interpretation (MacWhinney, Bates & Kliegl, 1984). German native speakers make use of an animacy heuristic in transitive sentences in which arguments compete for the event instigator role if deterministic morphological cues to interpretation (case/agreement) are unavailable (Frisch & Schlesewsky, 2001). Animacy violations are especially aggravating in such ambiguous contexts and lead to increased N400 amplitudes in response to assumed subjects/actors. In the present study, in contrast, the absence of competition for the actor role reduced reliance on the animacy heuristic and thus potentially further attenuated the animacy mismatch reflected in the N400.

Genre categorization did not modulate the semantic N400 effect but it affected the latency and scalp distribution of the ensuing P600, suggesting both temporal and functional differentiation in controlled conflict resolution. It seems tempting to conclude from these differences that the earlier P600 in the neutral group is a marker of well-formedness evaluation grounded in the default expectation of semantic congruency, and that the later P600 in the poetry group reflects pragmatic processing triggered by the knowledge that seemingly incongruent statements may in fact be highly significant in poetic contexts. This explanation is in line with the idea that literary genres are conventionally associated with specific

interpretive strategies (Culler, 1975; Schauber & Spolsky, 1986) and it accounts parsimoniously for both the ERP pattern and the results of the meaningfulness rating. However, there are reasons to believe that the P600 effects and the processes that they reflect may, in fact, not be that different. First, they are remarkably similar in terms of their morphology, which suggests merely a delay of the same process. Second, the topography of the effect in the neutral group is unusual and potentially misleading since it quite likely reflects component overlap with the posterior N400. If so, the poetry condition would differ neither in terms of top-down control on semantic access nor necessarily in terms of the processes involved in controlled conflict resolution but merely in terms of when conflict resolution is initiated. Thus, it is not entirely clear whether the late positivity effects merely index default vs. delayed processing, or whether they are indices of functionally distinct processing stages. In any case, the instruction-induced differences in late semantic evaluation match the effect of genre on the meaningfulness of incongruent sentences observed in offline evaluations, and thus corroborate the notion that later stages of comprehension are affected by the schematic genre representations that experienced readers have acquired.

To conclude, our findings disfavour the notion that genre categorization results in prosodically and semantically mediated lexical access if reading rate is controlled for; it remains an open question, though, whether genre affects the predictability component of the N400. Instead, our results indicate that literary and non-literary real-time comprehension differ in the controlled evaluation of semantic congruency and meaningfulness. Morphological licenses in poetry have no effect on meaningfulness but effectively induce subtle metrical differences which modulate readers' implicit prosodic expectations; genre awareness moderates the adverse effects of these formal poetic licenses.

[T]ext types function as programs of expectations that control and limit the task of processing actual texts.

Robert de Beaugrande (1978: 6)

2

Poetry-specific sentence processing B: Oscillatory dynamics

2.1 Genre expectations and the anticipatory gating of information

Schematic representations of prior knowledge are essential for comprehension (Bartlett, 1932; Ghosh & Gilboa, 2014; Hintzman, 1986; Kintsch, 1998; Minsky, 1975; Rumelhart, 1975, 1977; Rumelhart & Ortony, 1977; Schank & Abelson, 1975, 1977; ; Sparks and Rapp, 2010; Spiro, 1977). Since the comprehension of written discourse in experienced readers involves prior knowledge about previously encountered texts, some models of discourse comprehension (e.g., Graesser, 1981; Meutsch, 1987; van Dijk & Kintsch, 1983; Zwaan, 1993) posit that text types and genres also correspond to *memory schemata*, i.e., to adaptable associative networks of generic information that emerge from multiple episodes of past reading experiences.

Text type schemata promote and serve strategic processing. They are supposed to comprise (a) generic information at all text levels: expectable lexicogrammatical features, canonical text structures and content, and pragmatic information like the intended purposes and affordances of a text type, and (b) the appropriate processing strategies and routines for dealing with the expected input in accordance with one's goals (e.g., Beaugrande, 1978; Meutsch, 1986; Paltridge, 1995; Rumelhart, 1980; Schmalhofer and Glavanov, 1986; van Dijk & Kintsch, 1983; Verdaasdonk, 1982; Viehoff, 1995; Zwaan, 1993). By generating a priori expectations about the purpose and structure of unknown texts and about the relative informativity of their components, text type schemata allow readers to selectively attend to information they assume to be relevant for their goals, and to disregard what they expect to be less important or even disruptive.

Whereas text type effects on reading behavior and text memory are relatively well-established (e.g., Fischer et al., 2003; Hanauer, 1998a; Koops van't Jagt et al., 2014; Tillmann & Dowling, 2007; Zwaan, 1993), neurophysiological investigation of genre-specific reading remains sparse (but see Altmann, Bohrn, Lubrich, Menninghaus, & Jacobs, 2014; O'Sullivan, Davis, Billington, Gonzalez-Diaz & Corcoran, 2015; Schumacher & Avrutin, 2011). This seems particularly surprising if one considers that the influential predictive coding hypothesis regarding the human brain (Friston, 2005, 2012; Friston & Kiebel, 2009; Gregory,

1980) holds a clearly compatible view, claiming also that perceiving sensory events, including reading, also depends on one's mind-set, i.e., on prior knowledge and context-specific expectations.

But how might such genre-specific expectations be implemented in the human brain? It has been claimed that, for simple and complex tasks alike, efficient execution depends on the increased engagement of some cortical regions and the controlled inhibition of others (Jensen & Mazaheri, 2010), with alpha oscillations (8-12 Hz) in the human electroencephalogram (EEG) being a potential correlate of such top-down regulation. More precisely, alpha activity is assumed to reflect the functional inhibition of task-irrelevant neuronal structures, thereby enhancing access to stored and (assumedly) task-relevant information (Klimesch, Fellinger, & Freunberger, 2011). It is furthermore believed to be actively involved in the "anticipatory gating of information" (Foxe & Snyder, 2011), i.e., in biasing attention when a stimulus is expected but has not yet been perceived and processed. There is mounting evidence for this view as, across perceptual modalities, pre-stimulus modulations of oscillatory alpha activity have been reported in experimental tasks that enforce anticipatory selective attention (Del Percio et al., 2006; Foxe & Snyder, 2011; May et al., 2012; Thorpe, D'Zmura, & Srinivasan, 2012; Thut, Nietzel, Brandt, & Pascual-Leone, 2006)). In the spatial cueing paradigm, for example, visual attention is directed towards specific target locations, and pre-stimulus alpha power reflects the selective/lateralized modulation of spatial attention that correlates with enhanced behavioral performance (e.g., Händel, Haarmeier, & Jensen, 2011; Kelly, Lalor, Reilly, & Foxe, 2006). The present study aims to show that attentional bias in sentence reading can be induced by cueing a genre schema, i.e., a complex cluster of culturally acquired knowledge.

2.2 Experiment 1c: Pre-stimulus alpha power in sentence reading

To test this hypothesis, we reanalyzed the EEG data of Experiment 1b, in which two differentially instructed groups of participants read short sentences and performed a probe recognition task after each trial. This experimental setup al-

lows to examine categorization-dependent processing differences under identical input conditions and task demands. In Experiment 1b, it enabled us to test whether genre categorization modulates the processing of morpho-phonological alternations and semantic incongruities. For the present purpose, it allows to investigate potential categorization-dependent differences in anticipatory attentional states. Assuming that alpha power indexes such increased focal attention (Benedek, Schickel, Jauk, Fink, & Neubauer, 2014), we predicted that, if genre categorization indeed induces anticipatory regulation of cortical engagement, grounded in genre expectations and independent of the behavioral task, then we should observe higher average alpha power for the poetry group than for the neutral group.

Experiments 1a and 1b revealed interaction effects of instruction and linguistic variables in behavioral performance, online processing, and offline sentence evaluations, but we additionally observed an unexpected ERP main effect of instruction at the sentence-initial constituent: amplitudes of an early right-anterior negativity (N300) were higher for the poetry group than for the neutral group. A similar but less pronounced effect was present at the sentence-medial subject noun phrase. Despite considerable inter-item variability in the sentence-initial phrase, we observed a clear and narrow N300 peak whose amplitude was insensitive to the morpho-phonological manipulation at this constituent. Based on these observations, we reasoned that it might be related to instruction-induced attentional modulations *prior to* reading rather than to stimulus-dependent differences *during* reading. Therefore, a secondary objective of the present study was to test the hypothesis that the N300 relates to potential genre-induced pre-stimulus differences in top-down control of anticipatory brain activity. If this processing effect is indeed systematically linked to the anticipatory gating of information, we should find a correlation of pre-stimulus alpha activity and N300 amplitude.

2.2.1 Methods

The participants, procedure, and EEG recording are described in detail in section 1.2.1. In order to allow estimating individual alpha activity, we recorded 2-3 minutes of resting-state EEG (eyes closed) before and after the main experiment. Crucially, analyses of resting state EEG activity revealed no between-group differences in terms of individual alpha power and individual alpha peak latencies (two-sample t-tests: alpha frequency: $t(46) = 0.23$, $p = .81$; alpha power: $t(46) = 0.18$, $p = .86$). We assume therefore that potential pre-stimulus alpha differences between groups are instruction-induced rather than participant-specific.

EEG preprocessing

EEG data were pre-processed with *MATLAB*-based *FieldTrip* routines (Oostenveld, Fries, Maris, & Schoffelen, 2011). We determined individual alpha peak frequencies from the resting state EEG and conducted time-frequency as well as event-related potential (ERP) analyses.

Estimation of individual alpha peaks: Participant-specific alpha frequencies were estimated from resting state EEG recorded before and after the main experiment; estimation followed the approach by van Albada and Robinson (2013). First, we constrained the analysis to a time window of 90 s during which the participants' eyes were closed, resulting in a discernible alpha-power increase (Figure 2.1A). The epoch of 90 s was subsequently filtered between 0.3 and 30 Hz with a 4th order butter-worth band-pass filter. We then selected epochs of 5 s from 1000 random positions within the 90 s epochs. If the peak-to-peak amplitude within selected epochs did not exceed 150 μV, epochs underwent multiplication with a Hanning window and a subsequent Fast-Fourier Transformation (FFT). Power values as squared absolute part of the resulting complex numbers were averaged across the 1000 random epoch samples (Figure 2.1B, left panel). In the resulting power spectra, the frequency and power dimensions were $\log_{10}$-transformed (Figure 2.1B, middle panel), and power values were

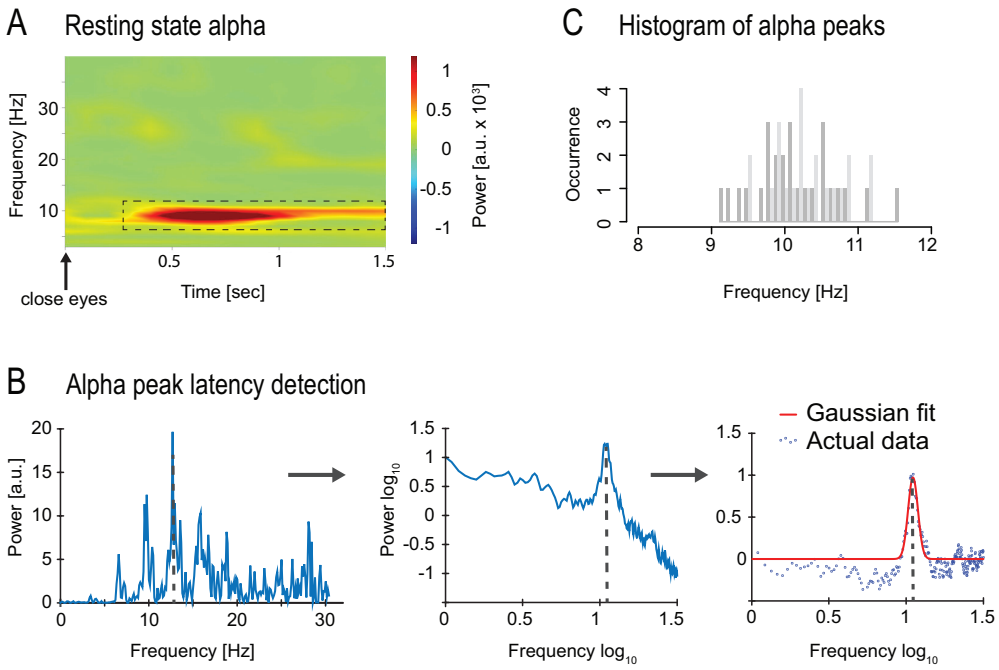

Figure 2.1: Exp. 1c: Estimation of individual alpha peaks from resting state EEG activity. From overall 90s-epochs (corrected for eye-related artifacts) we drew 5s-epochs at 1000 random latencies, uniformly distributed over the 90s-epochs. The beginning of one 90s-epoch from a representative participant is illustrated in A, showing increased alpha after eye closure. The 5s-epochs were individually Fourier-transformed and averaged, yielding a power spectrum as displayed in B (left). Power-spectra were $\log_{10}$-transformed and de-trended (B middle), before we fitted a Gaussian function to the data in order to estimate the frequency location of the participants' alpha peak (B right). Detected individual alpha-peaks ranged from 9 to 12 Hz (C).

detrended, using a linear regression approach. Subsequently, we fitted a one-peak Gaussian function to the transformed power spectra between 5 and 15 Hz from the parietal POz electrode using the FIT function within *MATLAB*. Frequency range and electrode selection was informed by previous results on alpha frequency determination in humans (e.g. Klimesch, 2012). If visual inspection of the resulting fits revealed distinct peak-fits (exemplified in Figure 2.1B, right panel), the corresponding frequency-value of the peak was determined. In 46 of the 48 participants in the main experiment, the procedure yielded estimable

alpha peaks from the resting-state EEG before the experiment. For the remaining 2 participants, alpha peaks were estimated from the resting-state EEG after the experiment. Estimated peaks ranged from 9 to 12 Hz, with a mean value of 10.2 ± 0.53 Hz (Figure 2.1C).

Regions of interest: Participant-specific alpha peak frequencies (rounded to the next integer) were then used in order to define spectro-temporal regions of interest (ROIs). These were based on bins of participant-specific alpha frequency ± 1 Hz in the frequency domain, and of two pre-stimulus windows in the time domain: one between -1.25 and -0.9 sec and one between -0.15 and +0.1 sec. Single-trial power values from the time-frequency analysis of the main experiment were extracted from these ROIs and averaged across all electrodes.

EEG preprocessing: The time-frequency and ERP analyses involved epoching the continuous EEG signal in 3-sec time bins with 1500 ms before and 1500 ms after the onset of the sentence-initial constituent (i.e., the PP). Epochs underwent automatic artifact rejection during which muscle and EOG artifacts (based on eye channels) were identified. If epochs contained such artifacts or exceeded peak-to-peak amplitudes of more than 150 μV, they were discarded from further analysis (14% data loss). Cleaned data were then referenced to the average of all channels and filtered between 0.3 and 40 Hz, using a 4th order zero-phase Butterworth band pass filter.

Time-frequency analyses: EEG data were re-sampled to 250 Hz without baseline-correction and de-trended using a linear regression approach before entering time-frequency analyses. Single-trial time-frequency analyses were based on Morlet-wavelets whose length (in cycles) depended on the frequency bin in the spectral domain. Bins reached from 1 to 30 Hz in steps of 1 Hz, and wavelet-cycles increase from 1 cycle at 1 Hz to 8 cycles at 30 Hz. In the temporal domain, the analysis window had a duration of 12 ms. Resulting complex values, centered on the wavelets, were obtained for each time-frequency bin, each channel and each trial. Power was calculated as squared absolute value of the complex values.

ERP analyses: Mean amplitude values of a 200 ms pre-stimulus window was subtracted from the epochs. ERP analyses focused on the N300 responses to

the sentence-initial PP. A 50-ms time window was centered at the negative peak (N300) between 250 and 350 ms post PP onset in the grand-average ERPs of all participants (318 ms ± 25 ms). For this latency range, we calculated single-trial averages at the right frontal electrode (F4) that showed the strongest N300 effect (see below).

Data analysis

In a first step, we tested for the fixed main effect of Instruction (Poetry vs. Neutral) on log-transformed single-trial power values from the two pre-stimulus windows using linear mixed-effects regression. For subsequent analyses that included pre-stimulus alpha power as predictor, we used values from the late time window immediately preceding stimulus onset. We then tested for fixed main effects and the interaction effect of Instruction and pre-stimulus alpha power on (a) N300 amplitudes (in μV), and (b) on log-transformed response latencies (in ms) of correctly answered trials, again using linear mixed-effects regression. All models contained random effects for both participants and items (Baayen, Davidson, & Bates, 2008). Whenever significant results indicated a relation between continuous variables, we tested the strength of this association using Pearson's Product-Moment Correlation.

Analyses were carried out in *R* (R Development Core Team, 2016) and based on the *lme4* package (Bates, Mächler, Bolker, & Walker, 2015). In order to facilitate interpretation and comparison, we report *F*- and *p*-value estimates based on the Satterthwaite approximation for degrees of freedom as implemented in the *lmerTest* package (Kuznetsova, Brockhoff, & Christensen, 2017).

2.2.2 Results

The time-frequency grand-average of the main experiment revealed differences in alpha power (9-12 Hz) between Instruction groups in two pre-stimulus baseline windows, following the onset and offset of the fixation asterisk that cued stimulus presentation (see Figure 2.2). The effect had a left-parietal topography

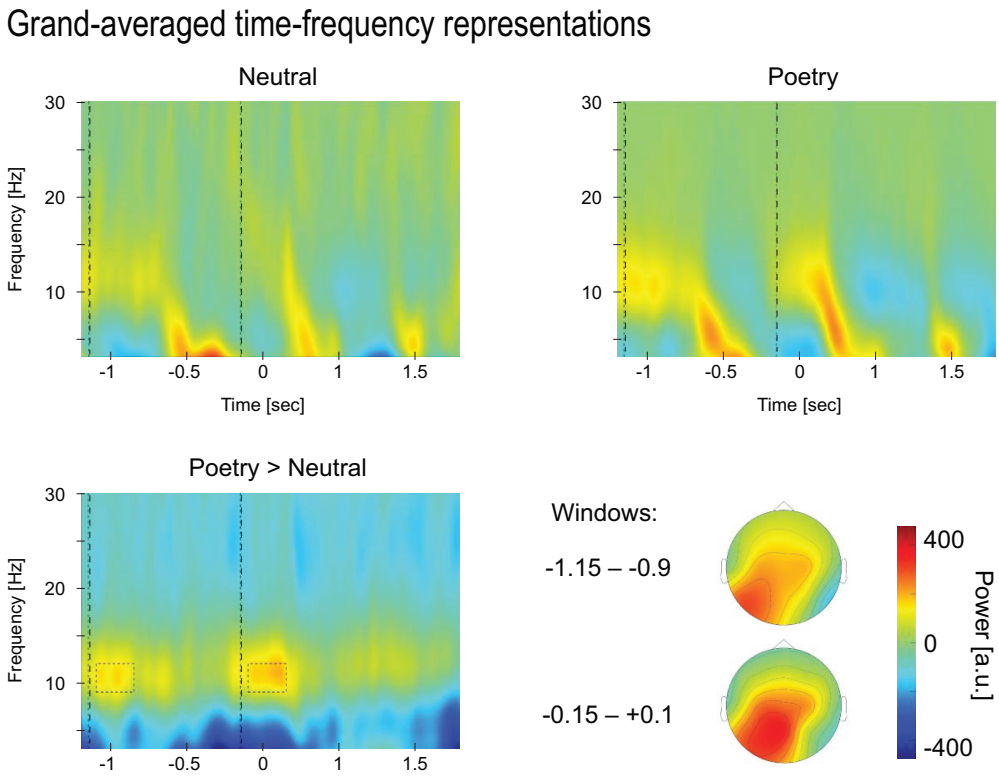

Figure 2.2: Results of the time-frequency analyses. Trials in the Poetry condition showed increased pre-stimulus alpha power compared to trials in the Neutral condition; the effect had a left-posterior scalp distribution. Vertical dotted lines indicate the onset and offset of the fixation cross prior to stimulus presentation (0 = stimulus onset).

in both windows and was more pronounced after cue offset. Statistical analysis corroborated these effects: Poetry instruction increased alpha power in both time windows (early: $F(1,4925) = 6.20$, $p = .013$; late: $F(1,4925) = 10.74$, $p = .001$; see Figure 2.3).

ERP responses to the sentence-initial PP differed in a right-lateralized fronto-central negativity (N300) peaking between 300 and 350 ms after PP onset (Figure 2.4A): a main effect of Instruction ($F(1,4923) = 7.14$, $p = .008$) reflected increased N300 amplitudes for the Poetry group vs. the Neutral group ($t(4923) = 2.67$, $p = .008$). There was an interaction of Instruction and alpha power

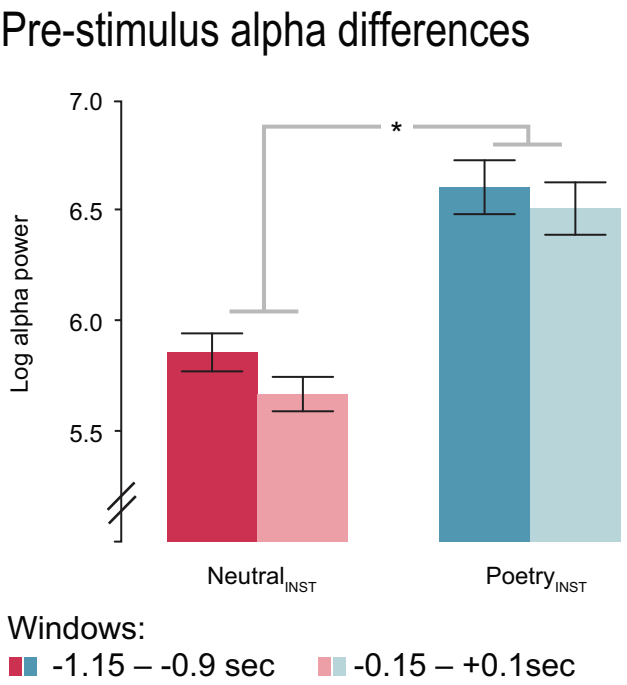

Figure 2.3: Average pre-stimulus alpha power in the Neutral (left) and Poetry (right) conditions. Bars indicate average alpha power in the early time window (blue) and in the late time window (red); error bars indicate the standard error of the mean; the asterisk indicates significance at $p < .05$.

($F(1,4923) = 15.40$, $p < .001$), which we resolved by Instruction. Pre-stimulus alpha power predicted early ERP responses in the Poetry group ($F(1,2571) = 33.90$, $p < .001$) but not in the Neutral group ($F(1,2349) = 0.31$, $p = .565$); for poetry readers, greater pre-stimulus alpha power corresponded to larger N300 amplitudes, i.e., more negative-going ERPs ($t(2571) = -5.76$, $p < .001$).

Testing for correlations between alpha power and N300 amplitudes at the trial level, we found a weak correlation in the Poetry condition ($r = -0.11$, $p < .001$) but not in the Neutral condition ($r = 0.01$, $p = .565$). At the participant-level, we found a moderate negative correlation between alpha power and N300 amplitudes for the Poetry group ($r = -0.37$, $p = .042$) but not for the Neutral group ($r = 0.12$, $p = .317$, see Figure 2.4B). Pre-stimulus alpha power further predicted performance in the post-stimulus probe task, as indexed by an effect of

alpha power on response latencies ($F(1,4916) = 17.93$, $p < .001$): Higher alpha power in the pre-stimulus period corresponded to faster reaction times in the probe task $t(4916) = -2.68$, $p = .007$). Neither the main effect of Instruction nor the interaction of Instruction and alpha power approached significance (Fs < 1, ps > 0.57). Correlation tests indicated a weak association between alpha power and reaction times at the trial level ($r = -0.05$, $p < .001$) and dissociation between the two variables at the participant level ($r = -0.11$, $p = .497$).

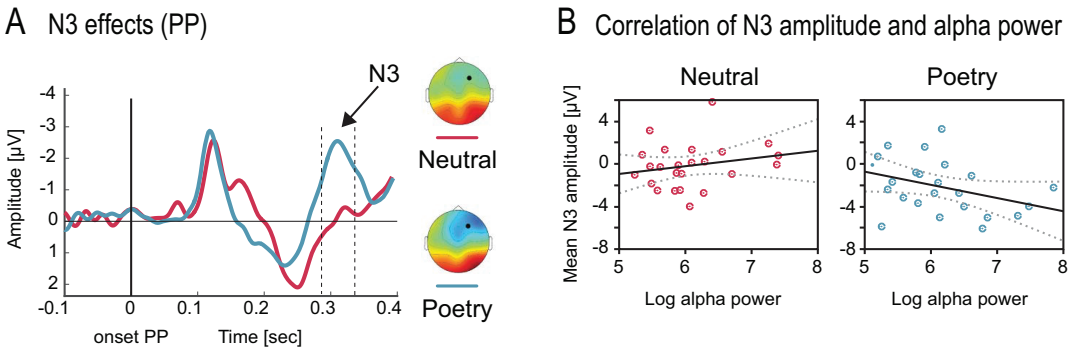

Figure 2.4: Exp. 1c: N300 effects and correlations with pre-stimulus alpha power. A. Grand-average ERPs time-locked to the beginning of the sentence-initial PP. Single-trial N300 amplitudes were obtained from a temporal region of interest (50 ms) centered around the overall negative peak between 250 and 350 ms, where the N300 effect showed a right-frontal topography with its center at the F4 electrode. Amplitudes for the poetry instruction were significantly larger than amplitudes for the neutral instruction. B. Illustration of the correlation between N300 amplitude and pre-stimulus alpha power (averaged per participant). The panels show the correlations of participant-specific averages of N300 amplitudes and log-transformed alpha power obtained from the late alpha ROI (-0.15 – +0.1 sec). Notably, the correlation was non-significant in the neutral condition, whereas N300 amplitudes and alpha power correlated negatively in the poetry condition.

2.3 Discussion and conclusion

Testing the hypothesis that text categorization (poetry vs. neutral) modulates attentional states prior to reading, we analyzed readers' pre-stimulus brain activity and early evoked responses in a dataset collected for an investigation of genre-specific sentence reading. In this study, two differentially instructed groups of

participants read identical sentences and performed the same behavioral task; the poetry group received instructions that categorized the stimuli as ”lines of verse” while the control group received neutral instructions that categorized the stimuli as ”sentences”. We expected that these groups differ systematically in terms of pre-stimulus alpha power, if text type attribution indeed modulates attentional states prior to reading. Our results confirmed this hypothesis: We observed that, compared to the neutral reading instruction (“sentence”), explicit genre attribution (“line of poetry”) was associated with (a) higher cue-related alpha power and with (b) increased N300 amplitudes in response to the sentence-initial syntactic constituent. We further found that (c) higher pre-stimulus alpha power was associated with reduced response latencies irrespective of the genre instruction.

To us, this pattern of results suggests that anticipatory attention allocation (as indexed by pre-stimulus alpha power) is driven by both genre expectations (text category predicts alpha power) and task requirements (pre-stimulus alpha power predicts response latencies), and that genre-induced attentional states prior to reading influence the early processing of verbal stimuli (alpha power predicts N300 amplitudes in the poetry but not in the control group). Crucially, the instruction groups did not differ in terms of individual alpha power and individual alpha peak latencies prior to the experiment. Moreover, participants in both groups read identical sentences and performed the same behavioral task so that cued differences in alpha activity most likely reflect instruction-induced attentional adjustments.

Earlier findings have already established a robust relation between pre-stimulus alpha and anticipatory selective visual and tactile attention (Del Percio et al., 2006; Foxe & Snyder, 2011; May et al., 2012; Thorpe et al., 2012; Thut et al., 2006); the present effect of text type attribution on pre-stimulus alpha power provides evidence for the genre-appropriate top-down control of cortical engagement in anticipation of a written verbal stimulus. More precisely, the poetry instruction seems to enhance the anticipatory gating of information (Foxe & Snyder, 2011), a mechanism assumed to be functionally dependent on pre-stimulus alpha power (Händel et al., 2011; Kelly et al., 2006). As this effect

was absent prior to the instruction and as it is observable already at the level of the single sentence, this finding further corroborates the notion of the genre schema as a control unit in reading whose activation precedes and modulates processing proper (Kintsch, 1998; Zwaan, 1993, 1996). This interpretation is supported by previous behavioral and electrophysiological results that show differential processing of identical textual input due to explicit text type attribution (Altmann et al., 2014; Schumacher & Avrutin, 2011; Zwaan, 1991, 1994), assumed to reflect cognitive strategies in accordance with readers' genre expectations (Zwaan, 1993) and processing goals (Vipond & Hunt, 1984). In this respect they resemble levels-of-processing effects in experimental tasks that selectively draw participants' attention to different linguistic levels of textual input (e.g. Bentin, Kutas, & Hillyard, 1993; Chwilla, Brown, & Hagoort, 1995; Craik & Robert, 1972). We assume that any genre schema provides a priori restrictions on the range of possible and permissible linguistic form, and thus creates general expectations regarding the characteristics of upcoming sensory input. However, whether the observed effect merely indexes the absence/presence of text categorization or whether it is indeed poetry-specific cannot be determined on the basis of the present results alone and remains open to future research.

Notably, this interpretation is also in line with prediction-based models of language processing that have a great potential to account for processing phenomena across modalities, timescales and complexity levels of perceptual input. For instance, recent neuroimaging data have provided evidence for expectancy-driven top-down modulation of low-level processing in audition (Chennu et al., 2013; Schröger, Marzecová, & SanMiguel, 2015; Sussman, Winkler, Huotilainen, Ritter, & Näätänen, 2002; Todorovic, Ede, Maris, & Lange, 2011), speech perception (Federmeier, 2007; Sohoglu, Peelle, Carlyon, & Davis, 2012) and visual perception (Egner, Monti, & Summerfield, 2010; Kok, Jehee, & de Lange, 2012; Summerfield & Egner, 2009), but also of higher-level processing in word and sentence comprehension (Bornkessel & Schlesewsky, 2006; Bornkessel-Schlesewsky & Schlesewsky, 2009; DeLong, Urbach, & Kutas, 2005; Federmeier & Kutas, 1999; Roehm, Bornkessel-Schlesewsky, Rösler, & Schlesewsky, 2007). Here, we observed that awareness of a text's genre modulates atten-

tional processes as reflected by alpha power, which seems to reflect that text-type schemata provide the necessary link between prior and future reading experiences that generates genre-specific expectations and triggers the attentional adjustments that we observed. We found further that increased cue-related alpha power correlated with more negative deflections between 300 and 350 ms (N300) only if a genre had been specified, suggesting that sensory processing is influenced by prior experience via the genre-induced differences in attentional states.

Given that the N300 proved insensitive to the morphological and semantic manipulations in the experiment, we consider it unlikely that it is a direct reflection of bottom-up input. Rather, we assume that it indexes an input-independent top-down contribution to processing that occurs as a reflex of the genre schema activation. But what exactly might these top-down contributions be that the N300 reflects? Previous studies support the idea that poetic texts are associated with highly emotional and affective properties (Kuehnast, Wagner, Wassiliwizky, Jacobsen, & Menninghaus, 2014; Menninghaus, Wagner, Wassiliwizky, Jacobsen, & Knoop, 2017). Connecting these findings to prior evidence that N300 amplitudes scale with the valence of visual stimuli (Carretié, Iglesias, & García, 1997; Carretié, Iglesias, García, & Ballesteros, 1997) and with the degree of emotionality in exclamations (Bostanov & Kotchoubey, 2004), one might assume that the N300 indicates the emotional-affective processing of visual verbal stimuli. If this interpretation is correct, we should observe modulations of N300 amplitudes in an experimental setup that investigates how poetry readers process emotional-affective words in visually presented sentences.

Another potential functional interpretation of the effect relates N300 amplitudes to the integration of orthographic and phonological/phonetic representations during reading (e.g., Bentin et al., 1993; Hasko, Bruder, Bartling, & Schulte-Körne, 2012; Spironelli & Angrilli, 2007). In this view, the N300 effect indexes that poetry readers indeed “attend to sound recurrences in the assumption that these are not random” (Beaugrande, 1978: 24). If this interpretation is correct, we should observe modulations of N300 amplitude in an experimental setup that contrasts how readers of poetry and prose process sound recurrences, e.g., al-

literation, in visually presented sentences. In any case, the exact functional specification of the observed N300 effect needs to be addressed in appropriately designed future experiments.

Finally, the observed effect of pre-stimulus alpha power on behavioral performance provides experiment-internal support for the idea that alpha activity is associated with participants' attentional states. Notably, this association was observable at the trial level but not at the participant level whereas the genre-dependent association between alpha activity and early evoked responses was stronger at the participant level than at the trial level. This pattern suggests that behavioral performance – although largely determined by stimulus properties and processing events – partially reflects the attentional state at the beginning of the trial prior to reading, a relation that is grounded in general attentional mechanisms shared by all readers. Genre-specific attentional states and processing routines, by contrast, seem to be reader-specific, which is consistent with the assumption that acquired genre schemata emerge from individual readers' prior reading experiences.

While genre effects could be accounted for theoretically in schema-based frameworks of text comprehension, it remained unclear how the processing system implements them. By focusing on how genre-related anticipatory responses are reflected in the EEG, we identify modulations of alpha power as a possible mechanism for implementing genre-specific reading strategies, reflecting readers' initial mindset and influencing the early processing of verbal stimuli. If the genre-specificity of the observed alpha effect is corroborated in future investigations, this method might offer a promising new approach to the study of text types and of how they are represented in the minds of readers.

We must rephrase our question and ask, "What does the reader do in these different kinds of reading?" Only when this is answered does it become realistic to ask what manner of texts may be especially conducive to aesthetic activities.

Louise Rosenblatt (1978: 23)

3

Readers as text processors and performers

3.1 Poetry- and prose-specific reading and text performance

Readers' conceptions of text types and genres allow them to strategically process and comprehend unfamiliar texts (e.g., van Dijk & Kintsch, 1983; Sparks & Rapp, 2010). Readers thus use past reading experience to optimise current reading performance, allocating processing resources in accordance with the exigencies and affordances of the text type, often their best guess at the text's structure and content prior to reading. Compared to non-literary comprehension strategies, *literary* modes of text comprehension re-rank the importance of surface form and situational information, and reduce the reading speed to deal with the inconsiderateness of literary texts (e.g., Hanauer, 1998a; Zwaan, 1991, 1994, 1996). Evidence for the differentiation of literary reading strategies is relatively sparse, though, and insufficient to inform a cognitive theory of literature and literary genres (e.g., Ryan, 1979; Steen, 1999). Moreover, previous studies typically focused on global indices of genre-specific reading performance, e.g., average reading rates, which are very useful to distinguish and characterize genre-specific reading strategies, but which are insufficient to capture differential dynamics during genre-specific reading.

The present study employed acoustic analyses and eye tracking during oral reading to address the differentiation of literary reading strategies, focusing on two broad categories of literary composition: literary prose and poetry. We examined whether these basic literary categories are associated with distinct reading strategies and differential phonetic recoding of written text, isolating top-down effects of genre categorization by keeping linguistic materials identical across genre conditions (e.g., Blohm, Menninghaus, & Schlesewsky, 2017; Bröggelwirth, 2007; Fant et al., 1991; Fischer, Carminati, Stabler, & Roberts, 2003; Koops van 't Jagt et al., 2014; Peskin, 2007; Zwaan, 1993). Contrary to prior investigations, we aimed to isolate dynamic aspects of genre-specific text processing, and to test two hypotheses regarding the genre-specific use of contextual information. Specifically, we examined whether poetry- and prose reading differ in how previously processed information affects (a) the phonetic realization of speech rhythm during oral reading and (b) word skipping during

text reading as an indicator of economic information gathering. Additional analyses of selected text-level measures sought to (c) replicate and refine the comparatively well-attested acoustic differences between poetry and prose performance, and to (d) extend the sparse contrastive eye-movement evidence for literary prose and poetry from silent to oral reading.

There is mounting cross-linguistic evidence for genre-specific processing strategies and performance styles (Barney, 1999; Blohm, Menninghaus, & Schlesewsky, 2017; Bröggelwirth, 2007; Byers, 1979; Fant, Kruckenberg, & Nord, 1991; Fischer et al., 2003; Hanauer, 1998a; Koops van 't Jagt, Hoeks, Dorleijn, & Hendriks, 2014; Schumacher & Avrutin, 2011; Wagner, 2012; Zwaan, 1993). Some previous studies have contrasted reading indices for different texts considered representative of the genres under investigation (e.g., Barney, 1999; Byers, 1979; Hanauer, 1998a), while others have opted for a same-text-different-genre paradigm in which only the experimenter-imposed text categorization but not the linguistic material differs between genre conditions .

The former approach allows selecting authentic texts with any number of genre-typical features, but makes it difficult to dissociate bottom-up effects of linguistic variables from top-down effects of text categorization. This line of research has revealed, for instance, that poetry is associated with slow reading rates and good verbatim recall (Hanauer, 1998a), and that memory for surface structure is more durable and accurate in poetry than in prose (Tillmann & Dowling, 2007). It has furthermore identified a set of melodic and rhythmical features that characterize 'poetic intonation' (Barney, 1999; Byers, 1979).

The same-text-different-genre approach, which is used in the present study, constrains the selection of experimental materials, since familiar texts and texts with salient and frequent genre cues (e.g., rhyme) could contradict and modulate the experimenter-imposed categorization (Fishelov, 1998; Hoffstaedter, 1987). Still, this approach may yield strong empirical evidence for genre-specific reading, because it minimizes the bottom-up influence of linguistic input and thus allows attributing observed effects to genre-induced adjustments of top-down control (Zwaan, 1993). With respect to literary prose and poetry, this line of

research has revealed adjustments of reading behavior (e.g., Fischer et al., 2003; Koops van 't Jagt et al., 2014; Peskin, 2007), interpretive strategies and text evaluation (e.g., Blohm et al., 2017; Gibbs, Kushner, & Mills, 1991; Peskin, 2007), as well as acoustic correlates of the respective performance styles (Bröggelwirth, 2007; Kruckenberg & Fant, 1993; Wagner, 2012).

Kruckenberg and Fant (1993) examined genre-specific performance styles and meter-specific speech rhythm in Swedish by analyzing a speaker's recitals of two poems as both poetry and prose. The global speech tempo was slower in poetry than in prose, and the durations of speech units (i.e., phonemes, syllables and prosodic feet) were longer. Stressed syllables were lengthened more than unstressed ones, which effects an increased S/W (strong/weak) contrast (i.e., metrical accent) in poetry performance. Finally, Kruckenberg and Fant noted that the relative duration of strong and weak syllables in a foot (i.e., the S/W ratio) distinguishes meters.

Bröggelwirth (2007) and Wagner (2012) investigated meter-specific prosodic prominence in a corpus of differently metered German poems that were first read as prose and then as poetry by speakers with varying expertise in text performance. The speakers were instructed in the poetry condition to attend to the metrical scheme of the poems and it was ensured that they had recognized the intended meter. The acoustic analyses revealed longer prosodic feet in poetry than in prose (as in English or Swedish) and relatively consistent S/W duration ratios of about 1.5, that were slightly lower for prose than for poetry and, contrary to the findings for Swedish poetry, similar for iambs and trochees.

Initial contrastive eye-movement evidence stems from a study by Fischer et al. (2003). Participants were presented with poems in both the original layout and in prose format (counterbalancing the order). When participants read the original poem versions, regressions were longer and more frequent, and progressive eye movements were shorter; average fixation durations were longer as well, but differences in reading rates and in the number of fixations did not reach statistical significance. Focusing on poetry's most characteristic formal feature, Koops van 't Jagt et al. (2014) demonstrated that lineation in poetry locally increases reading times and affects regressions at the words preceding

and following line breaks compared to prose versions where these regions occurred line-medially.

Taken together, the existing evidence supports the view that different genres are associated with distinct modes of gathering, processing and storing the information of a text, including specific conceptions of the appropriate phonetic realization. Across studies of silent and oral reading, poetry and prose seem to be most consistently distinguished by their defaults for reading and speech tempo (Byers, 1979; Hanauer, 1998a; Kowal, O'Connell, O'Brien, & Bryant, 1975; Kruckenberg & Fant, 1993; Peskin, 2007). Their diverging performance styles are well-attested and have been described in terms of distinctive acoustic correlates, some of which are supported by converging cross-linguistic evidence. Their (silent) reading modes also seem to differ in the reading speed defaults. Finer characterization in terms of distinctive eye-movement correlates remains elusive due to the scarcity of contrastive eye-movement evidence, and is further complicated by the absence of genre-distinctive global reading rates in the study that provided the most detailed characterization so far (Fischer et al., 2003). Finally, genre-specific reading has usually been characterized in terms of global adjustments as if it were static and stable throughout a text, and only few studies examined local effects that shed light on its dynamics (e.g., Koops van 't Jagt et al., 2014). It remains unclear, for instance, whether the greater rhythmicity (i.e., increased S/W contrasts) of poetry vs. prose performance reflects the genre-imposed profiling of speech rhythm or whether it hinges on readers' awareness of the metrical scheme that underlies the generation of the text.

3.2 Experiment 2: Poetry- vs. prose-specific oral reading

We combined eye tracking and speech recordings to investigate the performance styles and reading strategies associated with literary prose and poetry in experienced readers. To capture dynamic aspects of these adjustments, we tested two hypotheses that were guided by the general idea that reading strategies instantiate the functional constraints of the respective genres (van Dijk & Kintsch, 1983; Zwaan, 1993). According to this view, reading behavior should

reflect that discourse coherence and cohesion are the organizing principles of conventional (literary) prose, and that the organization of conventional poetic discourse highlights the actual form of the message, or may even be entirely dominated by formal constraints like meter and rhyme (Halliday & Hasan, 1976; Havránek, 1964; Jakobson, 1960; Mukařovský, 1964; Youmans, 1983).

Our first hypothesis concerns the genre-specific use of prosodic information, and derives from the aforementioned claim that poetry readers "have as part of the text-type frame the instructions to attend to sound recurrences in the assumption that these are not random" (Beaugrande, 1978: 24; see also Rosenblatt, 1978). If this claim is correct, we might expect that poetry readers attend to the recurrent prosodic patterns of meter, which is a prototypical and hence expectable feature of (German) poetry. Meter provides readers with almost continuous evidence of the patter, and if indeed they attend to recurrent prosodic configurations in the assumption that these are not random, they should recognize simple patterns of regularly alternating stressed and unstressed syllables (e.g., trochaic meter) after just two lines of a poem, i.e., after one immediate repetition. Once extracted, the metrical scheme should generate expectations that guide the further reading of the poem and lead readers to assign metrical accent to syllables in strong grid positions (Breen, 2018). Hence, we hypothesized that the availability of a regular prosodic context increases the rhythmicity of poetry (but not of prose) performance.

Our second hypothesis derives from a combination of known discourse effects on reading behavior and observations of genre-specific reading: Word skipping during reading is usually more frequent when contextual constraint and parafoveal preview are sufficient to recognize upcoming words without fixating them, thus warranting less careful navigation (e.g., Rayner, 1998). Previous contrastive eye-tracking results point to a similar top-down effect of genre: The reading strategy for poetry seems to be less "risky" than the strategy for prose (Fischer et al., 2003; Vančová, 2014) in the sense that it features shorter progressions and less word skipping (McGowan & Reichle, 2018; Rayner, Reichle, Stroud, Williams, & Pollatsek, 2006). Connecting these observations, we propose that only the prose (but not the poetry) strategy makes use of prior dis-

course context to optimize comprehension and shows more frequent word skipping in text-medial position.

We aimed to test these hypotheses locally, i.e., by investigating a predetermined critical region at which acoustic and behavioral effects of context and genre should become apparent, and for which our hypotheses could be translated into explicit and testable predictions based on existing evidence. Therefore, we decided to create short texts that contained sequences of monosyllabic function words, reasoning that they constitute the ideal testing ground for the proposed hypotheses due to their lexical and prosodic characteristics. Function words are closed-class words (determiners, prepositions, pronouns, etc.) that express more or less fixed "grammatical meaning" and that tend to be short and highly frequent; the ten most frequent words in English[1], for instance, are all monosyllabic function words (counting auxiliaries have and be) with an average length of 2.3 characters. Unlike content words, which obligatorily bear lexical stress, function words can be stressed or unstressed (English: Selkirk, 1996; German: Vogel, van de Vijver, Kotz, Kutscher, & Wagner, 2015), i.e., they are prosodically underspecified. Vogel et al. (2015) demonstrated that function words in German exhibit larger stress-level variance than content words, that their actually perceived and produced stress level depends on prosodic context, and that, where possible, even their syntactic position is co-determined by rhythmic constraints. During reading, function words show lower fixation rates (approx. 35%) than content words (approx. 85%); cf. (Carpenter & Just, 1983; Rayner, 1998; Rayner & Duffy, 1988), mainly due to their word length, as 2-3 letter words are only fixated around 25% of the time (Rayner, 1998; Rayner & McConkie, 1976). Based on these observations, we assumed that both the imposition of metrical accent and strategic word skipping should become apparent in our critical function word sequences.

To sum up, our combined investigation of poetry- and prose-specific reading and performance employed a same-text-different-genre paradigm that contrasted genres within readers while avoiding reader-internal text repetition (cf.

[1] According to the Corpus of Contemporary American English (Davies, 2008), which is available online at http://corpus.byu.edu/coca/.

Koops van 't Jagt et al., 2014). We expected to replicate the poetry-specific reduction of global speech tempo and reading speed, and to further characterize these effects in terms of distinctive features identified in earlier work or to provide additional evidence where the distinctiveness of features is disputed or unclear. Our text-level acoustic analyses focused on global speech rates, whose genre-distinctiveness has been challenged by Barney (1999), and on the degree to which they reflect independent contributions of speech unit lengthening (i.e., articulation) and speech pauses. Our text-level eye-movement analyses focused on global reading rates, whose genre-distinctiveness could not be confirmed by Fischer et al. (2003), and on the degree to which they reflect independent contributions of fixation durations and fixation frequency.

Presenting critical regions in either text-initial or -medial position allowed us to assess how contextual information alters the processing of identical linguistic input, and whether this influence is modulated by the reading strategies for literary prose and poetry. We expected to observe changes from text-initial to text-medial positions in terms of word skipping behavior and the phonetic realization of prosodic prominence, and to observe genre-specific dynamics in these changes.

3.2.1 Design and hypotheses

We used a 2 x 2 design that crossed genre (poetry vs. prose) and the text position (initial/medial) of critical regions (see next section) within participants and texts by means of a Latin square (cf. Koops van 't Jagt et al., 2014). The example in Table 3.1 illustrates how presenting identical texts in either a four-line stanza format (genre: poetry) or, removing two of the line breaks, in a two-line prose layout (genre: prose) kept linguistic variables constant across genre conditions and thus avoided potential confounds between text and genre (cf. Bröggelwirth, 2007; Fischer et al., 2003; Koops van 't Jagt et al., 2014).

We ensured that (a) the critical regions occurred line-initially in both genre conditions, and that (b) none of the line breaks necessitated syntactic re-analysis at the beginning of the next line (in most cases, line breaks corresponded to

Table 3.1: Example text in the four experimental conditions. Genre conditions (columns: prose/poetry) differed in terms of text layout and instruction. Position conditions (rows: initial/medial) reflected the order of the sentences. The critical and post-critical regions are underlined.

Prose – initial	**Poetry – initial**
Es hat tausend Jahre lang gedauert, bis man die Idee verstand. Oftmals irren selbst die allerklügsten unter allen klugen Köpfen.	Es hat tausend Jahre lang gedauert, Bis man die Idee verstand. Oftmals irren selbst die allerklügsten Unter allen klugen Köpfen.
Prose – medial	**Poetry – medial**
Oftmals irren selbst die allerklügsten unter allen klugen Köpfen. Es hat tausend Jahre lang gedauert, bis man die Idee verstand.	Oftmals irren selbst die allerklügsten Unter allen klugen Köpfen. Es hat tausend Jahre lang gedauert, Bis man die Idee verstand.

punctuation-marked clause boundaries). In line with genre conventions, the poetry layout differed from the prose layout in terms of line-initial capitalization (cf. Fischer et al., 2003) irrespective of word class, which otherwise determines sentence-internal capitalization in German orthography; note, however, that this did not affect the critical regions, as these featured sentence-initial capitalization in both genre conditions. In addition to the visual layout cue, participants received instructions that explicitly specified the genre of the texts (cf. Zwaan, 1993). The instructions and their English translations are provided in 3.2.

Table 3.2: English translations of the instruction texts for the Poetry and Prose blocks. Boldface highlights differences between instructions.

Prose instruction
In the following block, you are going to read 48 **excerpts from prose texts**. These **prose texts** originate from different **text** collections of mainly contemporary **authors**. Your task is to read these **prose texts** aloud.

Poetry instruction
In the following block, you are going to read 48 **stanzas from poetic texts**. These **stanzas** originate from different **poetry** collections of mainly contemporary **poets**. Your task is to read these **stanzas** aloud.

The text position of the critical region was manipulated by inverting the order of the two sentences within the text. Thus, we were able to examine the influence of prior prosodic context on participants' rhythmic performance of the critical region while keeping lexical and grammatical processing demands identical across position conditions.

We expected that the genre manipulation affects global eye-movement and acoustic measures of the reading tempo (poetry < prose) and the number and duration of fixations and speech pauses that participants make (poetry > prose). Based on Wagner (2012), we predicted that the local S/W ratio of critical feet approximates 1.5 more closely in poetry than in prose performance. If this rhythmicity effect reflects that poetry readers generally emphasize linguistic rhythm, we should observe a stable contrast across text positions (main effect of genre: 1.5 > poetry > prose). But if readers' awareness of the metrical scheme is a precondition for increased rhythmicity, then the contrast should arise only text-medially (interaction effect of genre and position: 1.5 > poetry medial > other conditions). Finally, if regular prosodic context reinforces the genre-dependent tendency to emphasize linguistic rhythm, then we should see an initial contrast that further diverges text-medially (main effect of genre + interaction: 1.5 > poetry medial > poetry initial > prose). We further predicted that text-medial words are skipped more frequently than text-initial ones (main effect of position: initial < medial), and that this tendency is absent in poetry reading (interaction effect of genre and position: prose medial > other conditions) or at least attenuated (main effect of position and interaction: initial > poetry medial > prose medial).

3.2.2 Materials and methods

Materials

We constructed 48 short critical texts, each consisting of two (typically bi-clausal) sentences. At least one of the sentences per text began with two monosyllabic function words (FWs), e.g., '*on the*', followed by a trochaic content word; see Table 3.3 for an example and the Supplementary materials (https://osf.io/ht6ar/) for the full set of stimuli. The texts were similar with respect to the number of words

(*M*=20.6, *SD*=3.6), characters (*M*=143.4, *SD*=9.8), and syllables (*M*=36.4, *SD*=2.5); cf. Table 3.4.

Table 3.3: Example of a critical text with English translation. The upper row indicates the underlying metrical pattern (S=strong; W=weak); critical and post-critical syllables are underlined; boldface indicates word stress; hyphens mark word-internal syllable boundaries.

S	W	**S**	W	**S**	W	**S**	W	**S**	W	**S**	W	**S**	W	**S**	W	**S**	W
Es	hat	**tau-**	send	**Jah-**	re	**lang**	ge-	**dau-**	ert,	bis	man	die	l-	**dee**	ver-	**stand**	
Oft-	mals	**ir-**	ren	**selbst**	die	**al-**	ler-	**klüg-**	sten	**un-**	ter	**al-**	len	**klu-**	gen	**Köp-**	fen
It has taken them a thousand years to understand the notion.																	
Often even the very wisest of the wise heads are mistaken.																	

The sentence-initial function word pairs (1st FW / 2nd FW) served as critical regions whose phonetic realization and word skipping rates we expected to differ between conditions. These pairs combined into the sentence-initial prosodic foot, and they conformed to one of four syntactic frames that allowed for sufficient lexical variation between items:

1. dummy pronoun + auxiliary (e.g., '*es hat*' = '*it has*')
2. subordinating conjunction + pronoun (e.g., '*als ich*' = '*when I*')
3. relative pronoun + determiner/pronoun (e.g., '*was ich*' = '*what I*')
4. preposition + determiner (e.g., '*auf der*' = '*on the*')

Each frame was represented by 12 critical items; the number of items per frame and condition was balanced for each participant. Post-critical content words were either free lexemes (44 items) or first constituents of compound words (4 items). They served (a) for comparison with the preceding critical foot and (b) to bias readers towards realizing the preceding critical region as a trochee to optimize rhythm (Vogel et al., 2015) and to parallelize adjacent feet (Kentner, 2015; Wiese, 2016; Wiese & Speyer, 2015).

All texts were presented in each of the four conditions; the resulting 192 critical text versions (48 items x 4 conditions) were distributed over four lists according to a Latin square (participants/texts/conditions). Each participant read

Table 3.4: Length parameters of critical texts and their critical and post-critical regions.

	Critical texts		Critical region				Post-critical region	
			1st FW		2nd FW		Content word	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
n words	20.6	2.6	1	–	1	–	1	–
n syllables	36.4	2.5	1	–	1	–	2	–
n characters	134.3	9.8	2.8	0.7	3.1	0.5	6.1	1.4

24 critical texts per genre condition, presented in separate blocks and with the appropriate layout. In addition to the constructed critical texts, each genre block additionally comprised 24 authentic texts (short prose excerpts or lyrical stanzas with various meters) that served as instruction-reinforcing fillers and were not analyzed. Each of the resulting eight genre-specific blocks was read by four participants.

Participants

We recruited 32 adult native speakers of German from the University of Frankfurt community (23 females; mean age = 25.4 years; range 18-39 years); all of them had normal or corrected-to-normal vision. The data of three further participants were excluded, as the removal of outliers and trials that contained speech errors would have resulted in less than 75% of the original observations in at least one condition; the recording session of one further participant was aborted due to ongoing calibration problems. All experimental procedures were approved by the Ethics Council of the Max Planck Society and were undertaken with no vulnerable populations involved and with the written informed consent of each participant.

Recording

Text presentation and audio recording were controlled via the OpenSource software *EyeTrack*[2]. We used a directional headset microphone (DPA Microphones

[2] *EyeTrack* is developed and provided by the Department of Psychological and Brain Sciences at UMass Amherst, MA: https://blogs.umass.edu/eyelab/software/.

A/S, Alleroed, Denmark) for voice recordings, sampling at 44.1kHz with a 16-bit resolution. The recording level was fine-tuned for each participant prior to the recording session.

Eye movements were registered with an EyeLink 1000 eye tracker (SR Research Ltd., Ontario, Canada) sampling at 1,000 Hz. Viewing was binocular, but only one eye was monitored (the right eye whenever technically possible). Participants were seated at a distance of approximately 60cm from the screen. Texts were aligned to the left edge of the display and presented in a black 28-point Courier font on a light-grey background; this presentation ensured that all texts could be displayed as intended and be read easily by the participants. Recording sessions began with a 9-point calibration of the eye tracker; a drift correction check was performed before each trial, and calibration was repeated after breaks and when deemed necessary by the experimenter.

Procedure

Prior to the main experiment, we informed participants about the experimental procedures and familiarized them with the technical equipment. Experimental sessions lasted about 45 minutes, comprising two blocks (poetry and prose; approx. 15 minutes each) separated by a short break (approx. 10 minutes). During the main experiment, participants were seated in a dimly lit and sound-attenuated recording booth while two experimenters (seated outside the recording booth) were documenting speech errors and monitoring the eye movement recording.

Participants received written instructions that specified the genre (prose or poetry) at the beginning of each block; the order of genres was counterbalanced across participants. Having made sure that the instructions had been read and understood, an experimenter fixed and checked the headset microphone and the experiment began with a 9-point calibration of the eye tracker and a short practice session of two trials per block. During the break, we removed the microphone and led participants into an adjoining room where they had a cup of water and filled out a brief questionnaire about their reading habits (approx. 5

minutes). Once they had finished the questionnaire, the experimental session was resumed with the second genre block, following the procedure described above.

Each trial began with the presentation of a small black square to the left of where the first word of the text was going to be displayed; text presentation and voice recording started once a valid fixation had been registered on this square. Participants were instructed to begin oral reading as soon as the text was displayed on the screen (to minimize conscious articulatory planning prior to articulation), and to press a button on a hand-held gamepad once they had finished reading the text.

Preprocessing and data analysis

We excluded all trials from the analysis that (a) contained speech errors (as documented during the experiment by one of the experimenters), or that (b) had late articulation onsets exceeding a threshold of three median absolute deviations (Leys, Ley, Klein, Bernard, & Licata, 2013), corresponding to 1.79 seconds post onset of the text display. On the basis of these criteria, 10.87% of the data points were excluded. The remaining observations were distributed evenly across genres ($\chi^2(1) = 0.09$, $p = .766$) and conditions ($\chi^2(3) = 2.81$, $p = .422$); articulation onsets of the remaining trials did not differ between genre conditions (Welch's two sample t-test: $t(1366.9) < 1$, $p = .694$).

Acoustic analysis

Voice recordings were pre-segmented using the software *MAUS*[3] (Kisler, Reichel, & Schiel, 2017). Subsequently, syllable boundaries in the critical region were checked and, if necessary, corrected manually, before we conducted the acoustic analyses using *Praat* (Boersma & Weenink, 2017). For each trial, we extracted the total speaking time per trial (i.e., the time from articulation onset to

[3] *MAUS* is developed and provided by the Institute of Phonetics and Speech Processing at LMU Munich, Germany:
https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface/WebMAUSGeneral

offset) and calculated participants' *speech rate* (total speaking time/number of syllables). We further extracted the number and lengths of the speech pauses and calculated the *total pause time* (i.e., the sum of all pause durations per trial) and the *average pause duration*. To dissociate effects of pauses and articulation on the speech rate, we calculated the *proportion of pauses* (total pause time/total speaking time), and the *articulation rate* by subtracting the total pause time from the total speaking time before dividing by the number of syllables per text.

For the analyses of local effects on text performance, we extracted the *foot durations* in the critical and post-critical regions, and the *duration*, *mean pitch*, and *mean intensity* for each syllable in these regions. We then calculated the *S/W ratio* per foot from the extracted values, assuming a trochaic rhythm (S+W) for both feet.

Eye-movement analysis

Analyses of global eye-movement measures were restricted to a subset of the measures reported in Fischer et al.'s (2003) investigation of genre-specific silent reading. Global *reading rates* index participants' reading performance, relating the amount of time readers spend collecting information from a text to the amount of information the text contains. Reading rates were calculated for each trial by dividing the summed fixation durations (i.e., excluding saccades) by the number of characters per text, yielding higher values when reading slows down. This global index was combined with the *number and the average duration of fixations* per trial to yield a slightly more nuanced characterization of how participants gathered the information from each text. Analyses of local effects were restricted to *skipping rates* for the critical FWs and the post-critical content words; we extracted for each trial whether the words in critical and post-critical regions were fixated or skipped. Fixations in the spaces between words were counted as fixations in the following region due to the asymmetric perceptual span (e.g., Ashby, Yang, Evans, & Rayner, 2012).

Statistical analysis

We analyzed trial-level data using mixed-effects regression with crossed random effects for participants and items (Baayen, 2008; Baayen, Davidson, & Bates, 2008); a forward-selection heuristic based on likelihood ratio tests (with an alpha level of 0.1) determined the parsimonious random effect structure for each model (Bates, Kliegl, Vasishth, & Baayen, 2015; Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017). We calculated Wald-type chi-square tests for the fixed effects of the final model.

In the analysis of the global data, we used linear mixed-effects regression to test for fixed main effects of genre on dependent eye-movement and acoustic measures. In the analysis of critical and post-critical regions, we tested for fixed main and interaction effects of genre and text position on (1a) absolute and relative foot durations and on (1b) absolute and relative syllable durations (in msec), intensity (in dB) and pitch (in Hz) using linear mixed-effects regression, as well as on (2) the likelihood of skipping (fixated/skipped) the critical words; we controlled the false discovery rate in multiple post-hoc comparisons (Benjamini & Hochberg, 1995). The skipping model further included word length as a covariate, the strongest predictor of word skipping (Brysbaert & Vitu, 1998; Rayner, 1998; Rayner & McConkie, 1976); this allowed us to control for the variance caused by word length differences between and within items. All analyses were carried out in *R* (Team, 2016) using the packages *car* (Fox & Weisberg, 2011), *lme4* (Bates, Mächler, Bolker, & Walker, 2015) and *lmerTest* (Kuznetsova, Brockhoff, & Christensen, 2017).

3.2.3 Results

Global differences between poetry and prose reading

We compared genre conditions with respect to a number of text-level variables identified in previous studies of poetry vs. prose reading and performance. Table 3.5 summarizes the observed values and the results of the statistical analyses.

Table 3.5: Global effects of genre (prose vs. poetry) on eye movements and articulation. Observed condition means and standard deviations (in parentheses) per genre, and results of the mixed-effects regression analyses: coefficient estimates (β) and standard errors (*SE*), 95-% confidence intervals ($CI_{95\%}$) and *t*-values.

	Prose	Poetry	β	*SE*	$CI_{95\%}$	*t*
Eye movements						
Reading rate (in msec/character)	57 (10)	60 (9)	3	±0.3	[2.2, 3.4]	8.81
Number of fixations (per word)	1.59 (0.22)	1.65 (0.21)	0.06	±0.01	[0.04, 0.08]	5.44
Average fixation duration (in msec)	237 (28)	240 (28)	3	±1	[1.2, 5.7]	2.98
Articulation						
Speech rate (in syllables/sec)	4.75 (0.62)	4.54 (0.57)	–0.21	±0.02	[–0.25, –0.18]	–13.10
Articulation rate (in syllables/sec)	5.28 (0.60)	5.09 (0.09)	–0.19	±0.02	[–0.23, –0.16]	–11.00
Proportion of pauses (in %)	10.3 (3.5)	11.0 (4.2)	0.8	±0.2	[0.4, 1.2]	4.29
Total pause time (in msec)	840 (383)	944 (475)	112	±18	[77, 148]	6.18
Number of pauses (per text)	3.0 (1.1)	3.4 (1.1)	0.4	±0.1	[0.3, 0.5]	6.53
Average pause duration (in msec)	302 (78)	286 (72)	–14	±6	[–27, –2]	–2.18

Global articulatory differences

Analyses of the acoustic data revealed that the global speech rate (incl. pauses) was about seven percent lower for poetry than for prose [$\chi^2(1) = 171.63$, $p < .001$]. This difference reflected that the poetry condition showed reduced articulation rates [–4%; $\chi^2(1) = 121.03$, $p < .001$] and more pause time per text [+12%; $\chi^2(1) = 38.14$, $p < .001$]; the proportion of speech pauses was greater in poetry than in prose [+7%; $\chi^2(1) = 18.42$, $p < .001$]. Closer examination of the pause length difference revealed that speech pauses were more frequent in the poetry condition [+13%; $\chi^2(1) = 42.60$, $p < .001$] while their average duration was shorter than in the prose condition [–5%; $\chi^2(1) = 5.04$, $p = .025$].

Global eye-movement differences

Statistical analyses revealed effects of genre on the global reading rate [+5%; $\chi^2(1) = 76.61$, $p < .001$], on the number of fixations per text [+4%; $\chi^2(1) = 33.03$, $p < .001$], and on average fixation durations [+1%; $\chi^2(1) = 9.04$, $p < .003$]; see Figure 3.1. Thus, participants read more slowly in the poetry condition than in the prose condition, making more and slightly longer fixations.

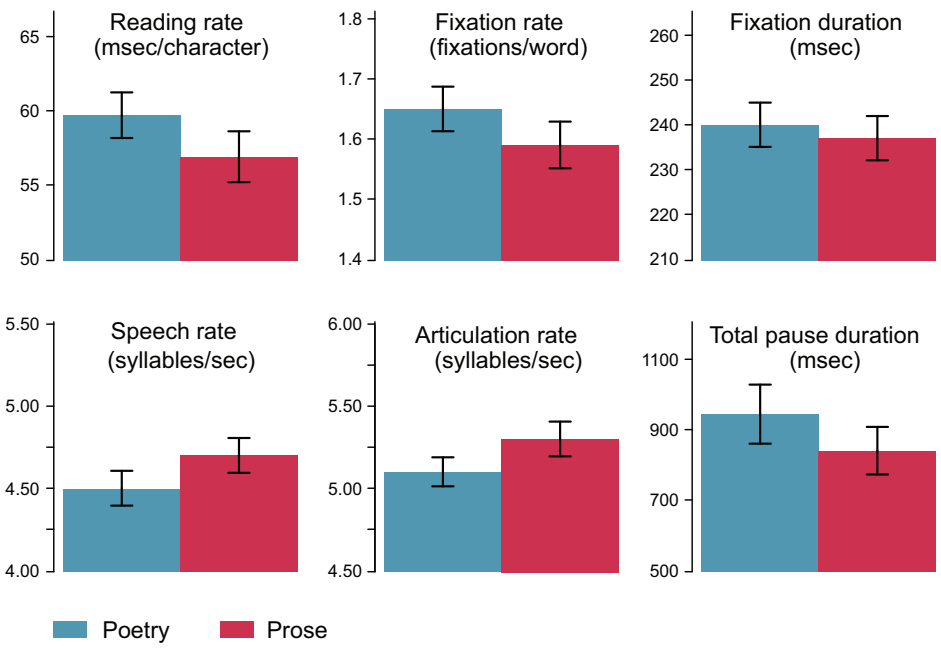

Figure 3.1: Exp. 2: Global effects of genre (poetry vs. prose) on eye movements (upper row) and acoustic measures (lower row) during oral text reading. Bars indicate observed condition means; error bars indicate the standard error of the mean.

Local effects of genre and text position

Local articulation results

We examined effects of genre and position on absolute and relative durations of critical and post-critical feet, and on foot-internal S/W ratios based on three acoustic correlates of syllable prominence (duration, intensity, and pitch).

Foot durations: We examined effects of genre and position on absolute and relative durations of critical and post-critical feet; the observed condition means and standard deviations are summarized in Table 3.6. Critical feet were affected by both genre [$\chi^2(1) = 8.09$, $p = .004$] and text position [$\chi^2(1) = 6.50$, $p = .011$]; their durations were longer in poetry than in prose [β = +9ms (±3ms)] and in text-initial than in text-medial position [β = +8ms (±3ms)]. Post-critical feet were likewise longer in the poetry condition [$\chi^2(1) = 12.31$, $p < .001$; β = +14ms

(±4ms)], but their durations did not vary between text positions [$\chi^2(1) = 2.17$, $p = 0.140$]. The relative duration of critical and post-critical feet was unaffected by genre and position [all ps > .36].

Table 3.6: Local articulation results: foot durations in milliseconds. Observed condition means and standard deviations (in parentheses).

	Critical foot		Postcritical foot		Relative foot duration	
	Prose	Poetry	Prose	Poetry	Prose	Poetry
text-initial	306 (48)	316 (43)	360 (53)	375 (53)	0.91 (0.11)	0.91 (±0.11)
text-medial	297 (41)	307 (49)	352 (51)	366 (44)	0.90 (0.08)	0.89 (±0.09)

S/W ratios in critical feet: Pitch- and intensity-based S/W ratios did not differ between conditions [all ps > .11]. Duration-based S/W ratios, in contrast, showed a main effect of genre [$\chi^2(1) = 4.11$, $p = .043$] that remained constant from text-initial to text-medial position [$\chi^2(1) = 0.09$, $p = .764$]; see Table 3.7. Detailed analysis of this effect revealed that strong syllables (i.e., 1st FW) were longer in poetry than in prose [$\chi^2(1) = 9.07$, $p = .003$; β = +8ms (±3ms)], and that their durations decreased from text-initial to text-medial position [$\chi^2(1) = 9.02$, $p = .003$; β = –8ms (±3ms)]. No difference between conditions was found for the durations of weak syllables, i.e., 2nd FW [all ps > .34]. Moreover, the pitch of FWs decreased uniformly from text-initial to text-medial position [1st FW: β = –9Hz (±3Hz), 2nd FW: β = –8Hz (±2Hz); both ps < .002].

S/W ratios – postcritical feet: Post-critical S/W ratios were unaffected by either genre or position [all ps > .21]. Syllable-level analyses disclosed an effect of genre on the durations of strong syllables [$\chi^2(1) = 9.87$, $p = .002$] which were longer in poetry than in prose [β = +8ms (±3ms)]. Contrary to the pattern observed for critical feet, weak syllables showed a similar trend [$\chi^2(1) = 2.94$, $p = .086$; β = +5ms (±3ms)]. As in critical feet, position affected the mean pitch of strong and weak syllables: the mean pitch of weak syllables was reliably lower text-medially than text-initially [$\chi^2(1) = 9.19$, $p = .002$; β = –5Hz (±2Hz)]; strong syllables showed a similar trend [$\chi^2(1) = 3.21$, $p = .073$; β = –3Hz (±2Hz)])].

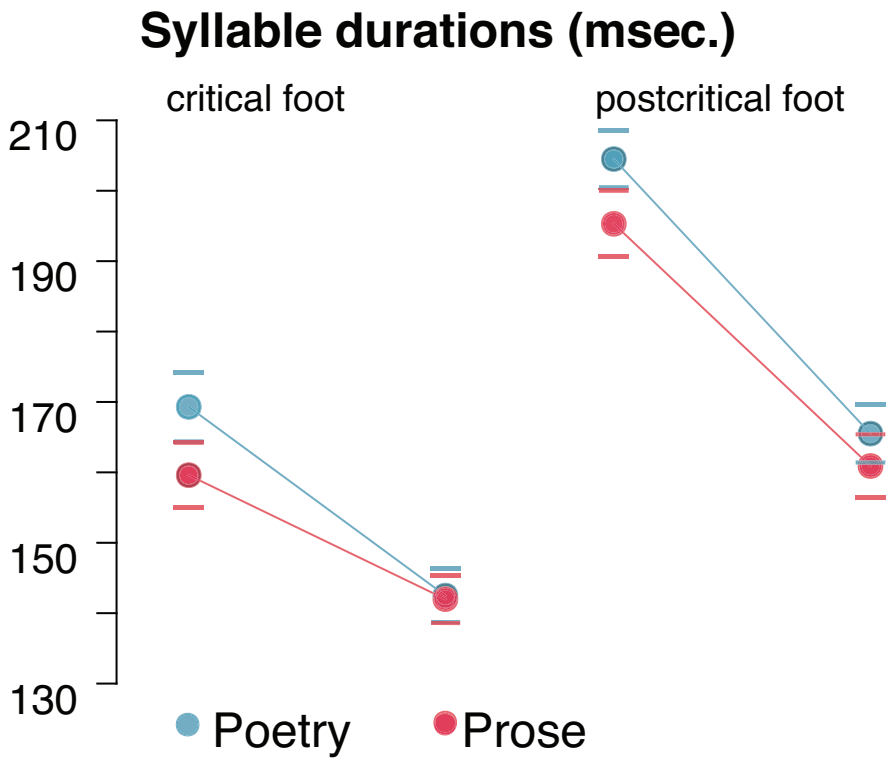

Figure 3.2: Exp. 2: Average syllable durations in critical (left) and post-critical (right) feet. Points represent condition means; error bars indicate the standard error of the mean; lines reflect foot-internal gradients of duration-based prosodic prominence, i.e., the durational difference between strong (S) and weak (W) syllables. The critical foot showed genre-induced modulations of the S/W ratio (prose < poetry), i.e., selective lengthening, whereas both syllables were lengthened in post-critical feet.

Local eye-movement results

We analyzed how likely our participants skipped the words in critical and post-critical regions. The summary of results in Table 3.8 reveals two general tendencies: For one, the post-critical content words were less likely to be skipped (approx. 1%–7%) than the relatively short function words (approx. 10%–50%). Moreover, word skipping increased from text-initial to text-medial position.

Main effects of text position confirmed that readers fixated text-medial words less often than text-initial ones. We observed the strongest effect at the 1st FW [$\chi^2(1) = 96.98$, $p < .001$; $z = 9.85$], but text-medial skipping extended to the 2nd

Table 3.7: Local articulation results: syllable durations, mean pitch, and mean intensity values. Observed condition means and standard deviations (in parentheses) for the syllables in the critical and post-critical regions.

Critical foot	1st FW (strong)		2nd FW (weak)		S/W ratio	
	Prose	Poetry	Prose	Poetry	Prose	Poetry
Duration (msec)						
text-initial	163 (31)	172 (29)	142 (23)	144 (23)	1.25 (0.19)	1.33 (0.29)
text-medial	156 (24)	166 (32)	141 (21)	141 (23)	1.22 (0.17)	1.29 (0.20)
Mean Pitch (Hz)						
text-initial	197 (51)	197 (47)	204 (49)	205 (47)	0.96 (0.06)	0.96 (0.05)
text-medial	189 (51)	189 (48)	196 (48)	196 (48)	0.96 (0.07)	0.96 (0.06)
Mean intensity (dB)						
text-initial	47 (3)	47 (3)	49 (3)	49 (2)	0.96 (0.05)	0.97 (0.04)
text-medial	47 (3)	47 (3)	49 (3)	49 (3)	0.97 (0.04)	0.97 (0.05)

Post-critical foot	1st syllable (strong)		2nd syllable (weak)		S/W ratio	
	Prose	Poetry	Prose	Poetry	Prose	Poetry
Duration (msec)						
text-initial	198 (34)	206 (35)	163 (28)	168 (28)	1.33 (0.20)	1.35 (0.24)
text-medial	193 (31)	203 (27)	159 (31)	163 (27)	1.34 (0.26)	1.39 (0.22)
Pitch (Hz)						
text-initial	205 (43)	205 (43)	211 (44)	213 (47)	0.98 (0.05)	0.97 (0.05)
text-medial	202 (46)	198 (44)	206 (44)	205 (44)	0.99 (0.05)	0.97 (0.06)
Intensity (dB)						
text-initial	50 (3)	50 (3)	49 (3)	48 (3)	1.04 (0.04)	1.04 (0.04)
text-medial	50 (2)	49 (3)	48 (3)	48 (3)	1.04 (0.05)	1.04 (0.04)

Table 3.8: Local eye-movement results. Average word skipping rates per condition and standard deviations (in parentheses).

	1st Function word		2nd Function word		Post-critical content word	
	Prose	Poetry	Prose	Poetry	Prose	Poetry
text-initial	14.36 (20.69)	11.80 (17.98)	26.13 (14.64)	24.48 (15.84)	1.71 (5.4)	2.44 (6.53)
text-medial	46.90 (30.91)	19.43 (25.20)	36.61 (25.01)	32.64 (22.12)	6.93 (19.55)	0.65 (2.62)

FW [$\chi^2(1) = 9.98$, $p < .001$; $z = 3.16$], and even to post-critical content words [$\chi^2(1) = 12.13$, $p < .001$; $z = 3.48$]. There were no main effects of genre [all ps $> .29$] but we observed interaction of genre and text position at the line-initial 1st FW [$\chi^2(1) = 18.20$, $p < .001$; $z = –4.27$] and the post-critical content word [$\chi^2(1) = 11.25$, $p < .001$; $z = –3.35$], but not at the intervening 2nd FW [$\chi^2(1) = 0.03$,

p = .873]; see Figure 3.3. These interactions reflected that readers skipped words in text-initial position equally often in both genres [1st FW $\chi^2(1) = 1.52$, p = .218]; content word [$\chi^2(1) = 0.42$, p = .517], whereas text-medial words were less likely to be skipped by poetry readers than by prose readers (1st FW: $\chi^2(1)$ = 64.31, $p < .001$; $z = -8.02$; content word: $\chi^2(1) = 11.82$, $p < .001$; $z = -3.44$). Greater word length decreased skipping rates only for the 1st FW [$\chi^2(1) = 9.26$, $p < .001$, $z = -3.04$], but neither for the 2nd FW [$\chi^2(1) = 0.78$, p = .377] nor for the post-critical content word [$\chi^2(1) = 0.11$, p = .741].

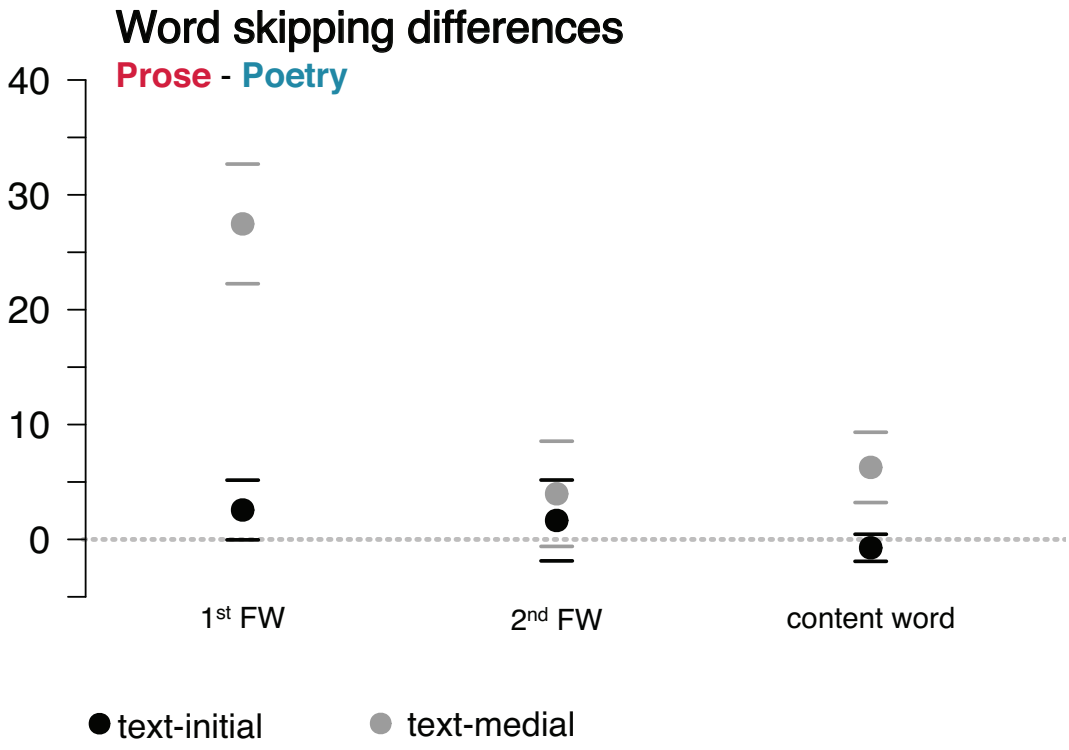

Figure 3.3: Exp. 2: Local word skipping differences (in %) between genres by text position (initial vs. medial) and critical word (1st FW, 2nd FW, content word); error bars indicate the standard error of the mean. Readers skipped text-medial function and content words more frequently in prose than in poetry.

3.3 Discussion and conclusion

Effects of genre and context on articulation

We observed slower global speech rates for poetry than for prose. This is in line with the list of melodic and rhythmical features that characterize poetic intonation as originally proposed by Byers (1979), but inconsistent with the revised list offered by Barney (1999). Non-expert readers quite consistently modulated their speech rate when orally reading different literary genres. Most (> 90%) of our participants reduced their speech rate for poetry vs. prose, which reflected similarly reduced articulation rates and more speech pauses per text. The articulation rate difference is consistent with previous cross-linguistic evidence (e.g., Swedish, English, and German) showing that readers realize speech units in poetry with longer durations than in prose (Bröggelwirth, 2007; Byers, 1979; Kruckenberg & Fant, 1993; Nord, Kruckenberg, & Fant, 1990; Wagner, 2012); in our study, we observed local effects on the durations of syllables and prosodic feet. The average articulation rate that we observed for poetry (5.1 syllables/second) matches the rate that Byers (1979: 369) reported for light verse rather than that for traditional poetry (4.8 syllables/second). Most likely, this primarily reflects that our "poems" were constructed to be consistent with contemporary prose. The finding that speech pauses occur more frequently in poetry than in prose also aligns with previous results (Barney, 1999; Byers, 1979) and seems mostly driven by the additional caesuras induced by the verse format (Kien & Kemp, 1994). Furthermore, we found an increased proportion of pause time per text in the poetry condition, which indicates that silent pauses added more reading time to the global speech rate effect than the additional articulation time per speech unit reflected in the reduced articulation rates.

The observed global differences support the view that even non-expert readers phonetically realize poetry and prose with distinct but (possibly) overlapping tempo ranges, modulating their articulation rate as well as the frequency and durations of speech pauses. This is consistent with the idea that genre-specific performance styles are distinguished by feature bundles rather than by single

distinctive features. The possible feature overlap between genres could thus account for Barney's (1999) observation of reduced speech rates in both a poetry recital and the performance of an advertisement by the same professional speaker. Greater expertise and increasing differentiation of genre conceptions seems to lead to increasingly pronounced performance styles and differentiated style repertoires (Funkhouser & O'Connell, 2013).

Our local analyses were restricted to the critical function word pairs and the lexical trochees in post-critical regions. We observed longer prosodic feet in poetry than in prose across regions and text positions, in line with prior investigations in English, Swedish and German; relative foot durations were equal across conditions. The genre-specific adjustment co-occurred with an independent effect of text position: the durations of critical feet (i.e., the function word pairs) decreased from initial to medial text position while the lexical trochees in post-critical regions remained unaffected, reflecting that function words are particularly prone to reduction (e.g., Vogel et al., 2015). As these effects of genre and position were of similar magnitude, we observed that poetry-specific lengthening roughly equalized the reduction of function word pairs in text-medial position.

Foot-internal prosodic prominence was measured by pitch-, intensity-, and duration-based ratios of strong and weak syllables. Pitch- and intensity-based S/W ratios approximated 1.0 across regions (i.e., including lexical trochees) and conditions and thus failed to capture relative prosodic prominence. The duration-based ratios, by contrast, ranged between 1.2 and 1.4 across regions and thus showed the expected prominence asymmetries. We observed greater S/W duration ratios in poetry than in prose, which indicates sharper prominence contrasts and thus more pronounced speech rhythm. While this finding generally aligns with earlier results (e.g., Wagner, 2012), we observed the genre-dependent highlighting of the prosodic profile only in function word pairs but not in lexical trochees. Syllable-based analyses revealed poetry-specific lengthening of strong syllables across critical and post-critical feet. Weak syllables, in contrast, were lengthened only in lexical trochees but not in pairs of monosyllabic function words. Only the lexical trochees showed S/W duration ratios con-

sistent with previous findings (approx. 3:2), whereas function word pairs had lower ratios (approx. 4:3) across conditions.

While corroborating the major findings of earlier studies, our results also reveal a more nuanced pattern of genre effects on the phonetic realization of prosodic prominence contrasts. Our findings replicate the poetry-specific lengthening of prosodic feet but refine previously observed genre effects at the syllable level. Only in lexical trochees did we find the previously reported pattern of absolute syllable durations, i.e., consistent syllable lengthening in poetry with greater effects on strong syllables. The previously reported pattern of relative syllable durations, i.e., increased rhythmicity during poetry performance (as indexed by reliable differences in S/W duration ratios), was observed only in the critical function word pairs but not in the lexical trochees of the postcritical region.

The differential genre effects on lexically fixed and prosodically underspecified feet suggest that monosyllabic function words play a special role not only in the rhythmic optimization of spontaneous speech (Vogel et al., 2015) but also of metered verse: they are not inherently stressed or unstressed, and it is this prosodic flexibility that allows creating artificial stress peaks where prose performance tends to show stress troughs. By contrast, polysyllabic content words feature inherent prominence asymmetries that are emphasized during poetry performance. Readers additionally lengthen stressed and, to a lesser degree, also unstressed syllables, but without necessarily affecting relative syllable durations.

The rhythmicity of poetry performance did not increase from initial to medial text position. Thus, our results do not support the idea that the poetry mode entails the early recognition of simple metrical schemes and their immediate application during text performance. The occurrence of such an effect would have provided strong evidence for the idea that increased attention to sound recurrences leads to the extraction of prosodic regularities which are then used to streamline text reading and performance. Its absence, however, is inconclusive with respect to both meter recognition and the genre-specific allocation of attention during reading. It might be the case that the poetry reading mode

simply does not entail increased attention to sound recurrences independent of their actual occurrence in a text. Alternatively, it might be that poetry readers do attend to prosodic patterns, but that either meter recognition takes longer than two lines of verse and presupposes the accumulation of more evidence (at least for non-expert readers like the ones in our experiment), or that readers do not apply the extracted metrical scheme during oral text reading, possibly with the intention to avoid an overly stylized performance. Finally, readers' increased attention might be restricted to perceptually salient sound recurrences at the sub-syllabic level, like rhyme or alliteration, rather than to the ongoing prosodic recurrence patterns that could just as well occur unnoticed over stretches of written prose or conversational speech (Schlüter, 2005). In any case, research using less indirect methods might be better suited to examine whether readers' attention to phonological structure is indeed subject to genre-induced top-down modulations.

A main effect of text position indicated that participants decreased the pitch of critical and post-critical syllables from text-initial to text-medial position irrespective of the genre, suggesting that literary prose and poetry converge in the general trend of the cross-sentential pitch contour. This decrease reliably affected prosodically underspecified function words and the weak syllables of post-critical trochees; syllables bearing lexical stress were only weakly affected. As this effect did not differ between genres, and as it did not alter the respective S/W ratios and the foot-internal prosodic prominence gradients, we assume that it reflects the falling pitch contour of the sentence pairs.

Effects of genre and context on eye movements

We observed greater global reading rates in poetry than in prose, i.e., readers spent more processing time per information unit when orally reading a poem vs. a prose excerpt. This slow-down was mainly driven by an increase in the number of fixations and, to a lesser degree, by their average duration. Since texts were identical across genre conditions, it is very unlikely that these global differences are due to the processing demands imposed by the texts themselves

(e.g., due to length, content, or style); they rather seem to reflect behavioral adaptations that are grounded in genre-adjusted processing goals and comprehension strategies (Meutsch, 1986; Peskin, 2007; Zwaan, 1993). But since task requirements (i.e., oral reading) were also equal across genre conditions, we assume that the global genre effects on eye movements reflect not only the comprehension-driven defaults of genre-specific reading behavior, but also the constraints of distinct performance styles on information flow.

The observed fixation-based reading patterns do not fully match the genre-specific profiles found earlier in reading poetry vs. prose for comprehension. A previous study (Fischer et al., 2003) also reported longer average fixation durations in poetry vs. prose whereas differences in reading rates and fixation frequency were, unlike in our study, not reliable. The source of this discrepancy is not entirely clear but it might be partially due to differences in material (authentic, mostly rhyming poetry vs. constructed non-rhyming quatrains in the present study). However, if it reflects differences in statistical power (Fischer et al.: 96 observations from 12 participants; the present study: >1200 observations from 32 participants), then the genre differences might in fact be quite similar across both studies. But there are further quantitative differences: fixation durations are slightly lower (5ms) in the present study whereas absolute global reading rates are considerably faster (25ms/char). The difference in fixation durations is expectable for oral vs. silent reading (e.g., Rayner, 1998). The discrepancy in reading rates may be partially due to the task: while readers typically proceed faster during reading for comprehension than during oral reading, it is only in the former task that performance profits from strategic re-reading as participants ensure that they have gathered the relevant information. Besides, some portion of the discrepancy in reading rates might reflect differences in the linguistic registers of the chosen text material: our constructed materials might have been easier to comprehend and navigate than the Romantic poetry that made up half of Fischer et al.'s (2003) materials.

First-pass skipping increased from text-initial to text-medial position for all critical and post-critical words. Line-initial words (=1st FW) were particularly affected by this increase, which seems to reflect the absence of prior linguistic

information for saccade planning in text-initial position. Moreover, fixations in text-medial regions were preceded by the return sweep between lines, whose landing site is usually six or more characters from the left line margin (Hofmeister, Heller, & Radach, 1999). However, since position effects occurred also line-internally (=at the 2nd FW and post-critical word), we assume that increased skipping rates in text-medial position also index that new information can be integrated into existing discourse representations, thus allowing for less careful reading than in initial text position.

This position effect was modulated by the genre. Text-initial word skipping did not differ between prose and poetry, but the genres diverged text-medially, where the skipping rate increase for line-initial function words (=1st FW) was greatly attenuated during poetry reading. The return sweep between lines might partially account for the genre difference at this position, since its landing site shifts rightward with increasing line length (Hofmeister et al., 1999), resulting in more word skipping of line-initial words in prose layout (approx. 60 characters per line) than in poetry format (approx. 30 characters per line). However, post-critical content words showed a similar pattern of effects: no difference between prose and poetry text-initially, and diverging skipping rates text-medially. But here, skipping rates increased only during prose reading, and remained stable or actually decreased during poetry reading. Thus, it seems that the genre asymmetry in line-initial position is partly a bottom-up effect driven by the differences in line lengths, and partly a top-down effect of readers' genre schemata that affects line-initial and line-internal words alike.

Limitations

We used constructed materials to increase experimental control, and results might look different if we had used authentic literary texts rather than genre hybrids. However, we believe to have succeeded in obtaining strong evidence for conventionalized processing routines that are activated by genre categorization independent of whether the texts are prototypical exemplars of their genre. We consider the presented evidence indicative of readers' schematic genre concep-

tions, and we assume that the observed behavioral correlates of genre categorization are even more pronounced for authentic texts. For if a text exhibits many distinctive or characteristic features of a literary genre or subgenre, it is likely to result in linguistic, semantic, and/or discourse representations that resonate with readers' genre conceptions (Hanauer, 1995, 1996). If bottom-up input thus further increases the activation level of the genre schema pre-activated by the categorization judgment, then the behavioral adjustments triggered by the genre should be reinforced as well.

Our results do not allow to assess how genre-specific reading routines and performance styles are modulated by expertise. Expertise is a key determinant of reader response to literary texts. It affects genre categorization (Hanauer, 1995) and comprehension strategies (Peskin, 2010), and it certainly affects performance styles (Funkhouser & O'Connell, 2013; Kowal et al., 1975). Here, we chose to sample from a student population so that we could presuppose enough experience with literary prose and poetry to ensure that participants had acquired distinct enough conceptions of literary prose and poetry. We believe that their mostly basic level of experience reflects an intermediate step towards the conceptual differentiation of literary genres that comes with greater expertise, and we assume that greater conceptual differentiation results in greater behavioral differentiation of genre- and subgenre-specific reading and text performance.

Conclusion

Replicating, extending and refining a number of previous results, our findings lend strong support to the idea that text reading is genre-specific, and that major literary genres are read and performed according to distinct routines. Previous and present results converge in the finding that literary prose is read faster than poetry, both silently and orally.

Prose seems to be read more economically as soon as the context provides sufficient top-down constraint to ensure successful comprehension; this tendency is clearly attenuated in poetry reading. When readers navigate through

literary prose, their progressive saccades are longer than during poetry reading, their regressions are less numerous and shorter, and their fixations are less frequent and of slightly reduced duration. The performance of literary prose is characterized by more rapid articulation than poetry performance: speech units are shorter and function words are reduced more frequently, there are fewer interruptions by speech pauses, and there is no additional emphasis on linguistic rhythm.

Such routines are straightforwardly derivable from approaches to discourse comprehension which posit that the control system in reading makes strategic use of text-type and genre schemata to optimize processing (e.g., van Dijk & Kintsch, 1983; Meutsch, 1986; Verdaasdonk, 1982; Viehoff, 1995; Zwaan, 1993). However, top-down modulations in the absence of linguistic differences are hard to accommodate within the Neurocognitive Poetics Model (Jacobs, 2015a, 2015b), a recent theoretical proposal that aims to relate stylistic theory to neuronal, affective-cognitive, and behavioral effects of literary comprehension. Arguably the most sophisticated formulation of the reception-related aspects of stylistic Foregrounding Theory, this dual-route model posits a fundamental distinction between explicitly processed foreground elements and implicitly processed background elements. While its current formulation does not (yet) aim to incorporate the neuronal, behavioral, and affective-cognitive effects of readers' genre categorizations, it could easily be extended to accommodate them by positing genre-specific background/foreground profiles that determine the appropriate processing defaults.

Whereas present and previous findings allow for a description of the behavioral correlates of genre categorization, the functional interpretation of the observed differences remains unclear. The co-occurrence of genre-specific text reading and performance patterns, however, suggests a functional link between the slow speech tempo of poetry performance and the reduced reading rates observed in silent reading. While prior results support the idea that behavioral correlates of genre-specific reading reflect different *comprehension strategies* (Hanauer, 1998a; Peskin, 1998, 2007), our findings support the notion that the demands of readers' (silent) *articulation strategies* put additional constraints on

eye movement control, and that the combined behavioral correlates of these strategies give rise to distinct eye movement signatures for literary prose and poetry.

Who forced [the poet] to compose poetry that will not be stable except by the commission of mistake? We have never seen or heard of a poet driven with a whip or a sword by a king or a person of authority to say that which is not permissible or allowed […] in the speech of others.

Ibn Fāris, (10th ct. AD)

4

Linking linguistic structure and poetic effect

4.1 Sentence judgments and the grammar of poetry

Competent speakers of a language have consistent intuitions about the grammatical status of verbal utterances. Even when confronted with novel expressions, they have no difficulty to distinguish the well-formed from "that which is not permissible or allowed". Such categorical grammaticality judgments are thought to reflect a speaker-hearer's linguistic competence, i.e., his implicit knowledge of his language, which is distinguished from the application of this knowledge in actual linguistic performance (Chomsky, 1965:4). Extra-linguistic constraints on the performance system, e.g., memory and attention, give rise to degrees of acceptability that reflect how easily comprehensible and how natural-sounding grammatical structures are. Native speakers' acceptability judgments reveal both categorical differences in grammaticality and degrees of acceptability, and thus provide an empirical basis for theories of linguistic competence and performance (Cowart, 1997; Schütze, 1996; Sprouse, 2013).

In this chapter, we will argue for the application of sentence judgment methods to the study of poetic competence[1] (Bierwisch, 1965; see also Culler, 1975), i.e., readers' implicit knowledge of literature in general and of poetry in particular. Sentence judgments provide a simple and objective way of testing hypotheses about the systematic covariation of linguistic form and poetic effect; they may thus support and constrain future theory formation in experimental poetics and, more generally, in the aesthetics of language and literature.

To illustrate this approach, we will present and discuss two experiments in which we contrasted grammatical (=acceptability) and literary-aesthetic (=poeticity) evaluations of single sentences. This combined collection of readers' grammatical and literary-aesthetic intuitions constitutes a methodological innovation that contributes to the inquiry into the relation between literary and non-literary language (e.g., Fabb, 2010; Fish, 1973; Mukařovský, 1964). We expect that it can reveal which linguistic features are attended to in poetry reading, and

[1] The term 'competence' is used in analogy to Chomsky's notion; this analogy has clear limitations and the two notions are not identical (see Abraham & Braunmüller, 1973; Pilkington, 2000).

how the evaluative criteria of this reading stance relate to the evaluative criteria of 'ordinary' grammar. Using this dual-judgment approach, we investigated readers' responses to two central characteristics of poetic language – deviant grammar and parallelistic sentence rhythms.

4.1.1 Deviation and parallelism in poetry

Poetry is widely considered a class of utterances that differs from most registers of the written standard, not least since poets across centuries and literary traditions have taken the liberty of tinkering with grammar when writing metered poetry (e.g., Aristotle, 1961; Lenerz, 1979; Sanni, 1993; Youmans, 1983). The genre is formally characterized by parallelism and licensed deviation from common language use (e.g., Fabb, 2010; Leech, 1969; Levin, 1962; Mukařovský, 1964). The term 'parallelism' as used here denotes all types of systematic patterning of linguistic elements, including phonological parallelism such as rhyme and meter (cf. Fabb, 1997; Hopkins, 1865; Leech, 1969; Menninghaus, Wagner, Wassiliwizky, Jacobsen & Knoop, 2017). The stylistic term 'deviation', on the other hand, covers both rule violations and deviations from the principles of non-literary language.

Parallelism and deviation are systematically related during both the production and the reception of poetic texts—in both literary practices, their effects are directly opposed. In poetry production, parallelistic schemata like meter and rhyme constitute self-imposed formal constraints on the selection and combination of linguistic elements (Bower & Bolton, 1969; Fabb, 2010; Idsardi & Raimy, 2005; Jakobson, 1960; Kintsch, 1998; Leech, 1969; Levin, 1962; Rice, 1997). The additional constraints of parallelism make writing more difficult and challenging, and poetically licensed deviation, i.e., the optional "commission of mistake", allows sidestepping some of these difficulties (e.g., Aristotle, 1961; Rice, 1997; Sanni, 1993; Youmans, 1983). In poetry reception, parallelistic sound patterning is beneficial in the sense that it facilitates early processing of verbal stimuli (Chen et al., 2016; Hoorn, 1996; Menninghaus, Bohrn, Altmann, Lubrich & Jacobs, 2014; Obermeier et al., 2016; Vaughan-Evans et al., 2016), assists their

memorization and recall (e.g., Rubin, 1995; Rubin, Ciobanu & Langston, 1997; Tillmann & Dowling, 2007; van Peer, 1990), and increases their aesthetic appeal and emotional impact (Kraxenberger & Menninghaus, 2016; Menninghaus et al., 2015; Menninghaus et al., 2017; Obermeier et al., 2013). Formal deviation, on the other hand, is linked to deautomatization in stylistic theory, i.e., to the impairment, extension and intensification of the reading process (e.g., Jacobs, 2015a, 2015b; Leech, 1969; Miall & Kuiken, 1994; Mukařovský, 1964; Šhklovsky, 1965; van Peer, 1986, 2007). This mechanism is believed to be responsible for many stylistic effects and there is empirical evidence to support this assumption (e.g., Chesnokova & van Peer, 2016; Hakemulder, 2004; Miall & Kuiken, 1994; van Peer, 1986). The degree of a text feature's deviance is co-determined by a number of factors including the diachronically evolving linguistic standard from which the feature or text deviates (Mukařovský, 1964), the linguistic norm of the text itself (Levin, 1965), and the linguistic and non-linguistic context (Enkvist, 1973; Riffaterre, 1960).

4.1.2 The present study

The present study constitutes an attempt to connect the study of deviant and parallelistic linguistic features and their poetic effects to the research tradition of experimental linguistics, where the use of sentence judgments is well-established. Following common practice in experimental syntax, we used acceptability judgments to estimate degrees of grammatical deviance (Sprouse, 2013). Building on earlier work in experimental poetics (Hanauer, 1995, 1996; Hoffstaedter, 1987; Martindale, 1973), we used poeticity judgments to estimate the poetic qualities of verbal stimuli. Although we further share Martindale's (1973) focus on the relationship between grammatical variation and poeticity, our study differs from its predecessors in some vital aspects (see section 4.2.4 for further discussion). One important difference is that collecting readers' intuitions about single sentences (rather than entire texts) allowed us to examine how poeticity relates to construction-specific grammatical deviance (cf. Martindale, 1973) and to prosodic parallelism in sentence rhythm. In other words, what we present is

a systematic study of how readers' intuitions about the poetic qualities of single sentences vary with the presence/absence of selected classes of grammatical features.

Experiment 3a seeks to establish that (1a) the conception of poetry shared by contemporary readers licenses some but not all degrees of grammatical deviance (cf. Fitzgerald, 2007; Lenerz, 1979; Martindale, 1973), and that (1b) the deviance of selected routine licenses contributes a distinctly poetic quality to the experience of sentence comprehension (cf. Hoffstaedter, 1987). Experiment 3b aims to provide initial support for the idea that (2) similarly poetic processing experiences are evoked by both syntactic and prosodic configurations that result from naturally occurring linguistic variation. As a methodological contribution, our study aims to demonstrate that sentence judgments offer a high enough resolution to allow linking poetic effects to specific linguistic features and to properties of sentence processing: Experiment 3a explores whether poeticity judgments reflect categorical and gradient differences in terms of genre-specific well-formedness; Experiment 3b examines whether they reveal similar evaluation criteria in the absence of clear genre cues, and whether they reflect structural features and processing effects of both syntax and (implicit) prosody.

4.2 Experiment 3a: Grammatical deviation in lyrical verse

In Experiment 3a, we focused on conventionally licensed grammatical deviation in traditional verse, i.e., on "the conformist aspect of poetic language" (Leech, 1969:15). These routine licenses of traditional poetry can be systematically accounted for by metrical (or rhyme) constraints (e.g., Youmans, 1983; Rice, 1997), but there are language-specific grammatical constraints that must not be overridden by the constraints of versification (Fitzgerald, 2007). We reasoned that these deviations would thus be an ideal testing ground for the idea that readers' genre-specific knowledge is reflected in categorical and gradient judgments of poeticity.

Following Leech (1969), we identified three classes of grammatical deviations from contemporary standard language use, all of which are routinely li-

censed in German poetry of the 18th and 19th century: syntactic inversions, morphological deviations, and prenominal genitives (see Materials section). For each of these, we selected original lines of poetry that were rhythmically regular. By carefully modifying these lines, we constructed alternative versions that were grammatically well-formed and canonical but rhythmically irregular. This allowed us to examine the effects of these deviations (and the resulting parallelism) without altering the lexical material or the meaning of the lines. Following Fitzgerald (2007), we further included a condition with deviations that are possible but virtually nonexistent in traditional German poetry. We will call this the control condition, as it allowed us to assess whether participants actually rely on their genre-specific knowledge, and whether they exhibit a limited degree of tolerance for metrically motivated grammatical deviation (Enkvist, 1973).

We expected that grammatical deviations would increase the perceived deviance of the verse lines and their disruptive potential during reading, and that this is reflected in low acceptability ratings. We further expected licensed deviations to increase the poeticity of the verse lines, reflecting (a) their status as genre markers and (b) the prosodic regularization that motivates and accompanies their occurrence.

4.2.1 Design and hypotheses

Our design crossed Grammar (Deviant vs. Canonical) and deviation Type (Syntax, Morphology, Genitive, Control) within participants, whereas Task (Acceptability vs. Poeticity) was a between-participants variable (n=40 per group). Overall, this mixed design yielded 5120 observations in sixteen experimental conditions (320 observations each); no observations were excluded.

We predicted lower average Acceptability ratings for Deviant vs. Canonical sentence variants regardless of the deviation Type (acceptability: Deviant < Canonical). Failure to find such an effect would indicate that our modifications were ineffective. If readers license only some but not all deviations, we should find that average Poeticity ratings for deviant Control sentences be lower than or equal to ratings for their Canonical alternatives (poeticity: $\text{Deviant}_{\text{Control}} \leq$

$\text{Canonical}_{\text{Control}}$). The absence of such an interaction effect would suggest that contemporary readers' poeticity judgments exclusively reflect deviance rather than genre-specific knowledge. We further expected higher average poeticity ratings for Deviant/parallelistic vs. Canonical/non-parallelistic sentences, if the presence of deviant grammar and parallelistic meter contributes a distinctly poetic quality to these lines and to the experience of reading them (poeticity: $\text{Deviant}_{\text{Critical}} > \text{Canonical}_{\text{Critical}}$). Failure to find such an effect would indicate that contemporary readers do not tolerate these deviations in poetry. Finally, we expected to find negative correlations between acceptability and poeticity that reflect how much the Type-specific deviance contributes to the perceived poeticity of the lines.

4.2.2 Materials and methods

For each deviation type, we selected 16 representative sentences from a corpus of German poetry dating from the late 18th to the early 20th century; each sentence corresponded to a single line or a non-rhyming couplet. Selected sentences were metrical and contained grammatical features that deviate from contemporary standard German (Grammar: deviant). We modified each of these sentences, constructing a non-deviant (Grammar: canonical) but metrically irregular version. As the control condition contained deviations that are, at best, marginal in German poetry, we could not simply select these items from a corpus but had to construct both versions based on original lines. Analogous to critical conditions, one version was grammatically deviant but metrical (Grammar: deviant), while the other was grammatically well-formed but unmetrical (Grammar: canonical). Deviant control items contained agreement errors (8 items, cf. English: '*he write_ a letter*') and unlicensed syntactic inversions similar to the ones that Fitzgerald (2007) described (8 items, cf. English: '*a he letter writes*'). Note that, across conditions, both versions usually contained further lexico-semantic, thematic and grammatical genre markers in addition to the manipulated variables; these included, for instance, other types of licensed grammatical deviations, genre-typical lexical items and word formations (e.g.,

'*nearward*' for '*nearer*'), and themes of song and dance, death and mourning, love and courting etc. See the Supplementary materials (https://osf.io/ht6ar/) for the full set of experimental stimuli.

Items were distributed over two lists, each containing all 64 items in either deviant (50%) or canonical (50%) versions. We did not include any distractor items, as the semantic/thematic differences between items, the differences between the four deviation types, and the counterbalancing of deviant vs. canonical sentences already resulted in considerable variation. For both lists, we prepared five pseudorandomized versions to reduce effects of item order; each of the resulting ten versions was judged by four participants per task.

Syntactic inversions (Type: Syntax)

Syntactic inversions arrange larger sentence constituents, i.e., syntactic phrases, in a way that does not conform to the rules and principles of 'ordinary' language (e.g., Leech, 1969; Youmans, 1983). By definition, they affect the order of meaningful elements and thus the incremental interpretation of a sentence (Cureton, 1980). Metrically, syntactic inversions correspond to rearrangements of larger prosodic constituents like phonological words and phrases; most inversions seem to be employed to satisfy the constraints of meter and/or rhyme (Fitzgerald, 2007; Youmans, 1983). Taking these features of poetic syntax as the common denominator, and neglecting a number of grammatical variables, we decided to lump together a variety of different syntactic inversions and treat them as one class for the present purpose.

In German, word order is partially free in the sense that it is, to some extent, determined by strict ordering constraints and partly governed by a number of potentially conflicting pragmatic and semantic linearization principles (e.g., Haider, 2010; Uszkoreit, 1986). The inversions we selected for this study were of both kinds. Some deviated from the order dictated by the strict, rule-like constraints and resulted in, or bordered on, ungrammaticality. These included, for instance, illicit verb placement or the occurrence of multiple constituents in the sentence-initial preverbal position of declarative main clauses (Haider, 2010:1;

but see e.g., Müller, 2003). Most inversions, however, violated linearization principles and resulted in very marked but grammatical structures, such as extrapositions (cf. Example 1a) or argument inversions (e.g., object-before-subject, noun-before-pronoun). We modified these inverted lines in accordance with the word order rules and linearization principles of German syntax, following Haider (2010) and Uszkoreit (1986); see Table 4.1 for an example.

Table 4.1: Example couplets for the Syntax conditions. Inverted constituents are underlined. Boldface indicates lexical stress and/or phrasal accent; asterisks indicate optional stress.

(4a) Deviant

Und als sie <u>kamen näherwärts</u>, begann der Grund zu zittern.
'*And as they <u>nearer came</u>, the ground began to shake*.'

Und	**als***	sie	**ka–**	men	**nä–**	her–	**wärts**,
be–	**gann**	der	**Grund**	zu	**zit–**	tern	

(4b) Canonical

Und als sie <u>näherwärts kamen</u>, begann der Grund zu zittern.
'*And as they <u>came nearer</u>, the ground began to shake*.'

Und	**als***	sie	**nä–**	her–	**wärts***	**ka–**	men,
be–	**gann**	der	**Grund**	zu	**zit–**	tern	

Morpho-phonological deviations (Type: Morphology)

Minor word-form alterations reduce or increase the number of unstressed syllables and they occur in many metered texts, e.g., lyrical poetry or contemporary pop songs. These word forms are usually infrequent or distorted variants of the canonical forms used in the current written standard. To make the lexical material fit the metrical grid, authors insert or delete reduced vowels, contract word stems, add or omit affixes, etc. In German poetry, examples include contractions (cf. English '*'twas*' vs. '*it was*'), omissions of otherwise obligatory affixes,

and uses of archaic inflections (cf. English '*he helpeth*' vs. '*he helps*'). Metrically, these morpho-phonological deviations are local adjustments of prosodic structure; they allow aligning stresses and strong metrical positions by avoiding stress clashes or lapses. As an illustration, consider the example in Table 4.2, in which historical but not contemporary inflections respect dactylic meter.

Table 4.2: Example couplets for the Morphology conditions. Affected words are underlined; boldface indicates lexical stress and/or phrasal accent; asterisks indicate optional stress.

(4a) Deviant: historical inflections
Das, was du da <u>siehest</u>, ist Totengeleit, und was du da <u>hörest</u>, sind Klagen.
'*That which thou <u>seest</u>'s a funeral train, and that which thou <u>hearest</u> are wailings*.'
Das, **was*** du da **sie**– hest, ist **To**– ten– ge– **leit**,
und **was** du da **hö**– rest, sind **Kla**– gen.
(4b) Canonical: contemporary inflections
Das, was du da <u>siehst</u>, ist Totengeleit, und was du da <u>hörst</u>, sind Klagen.
'*That which you <u>see</u> is a funeral train, and that which you <u>hear</u> are wailings*.'
Das, **was*** du da **siehst**, ist **To**– ten– ge– **leit**,
und **was** du da **hörst**, sind **Kla**– gen.

Prenominal genitives (Type: Genitive)

Genitive constructions prototypically express possession ('*Bob's book*') or part-whole relations ('*the top of the hill*'); grammatically, they relate the two nouns expressing the possessor and the possessed. Genitive constructions come in two types: prenominal ones and postnominal ones. In a prenominal genitive construction the possessor precedes the possessed entity (e.g., '*Jennifer's jacket*'); in a postnominal one the possessor follows (e.g., '*the back of the car*'). In German, the postnominal genitive construction is canonical for most noun classes (e.g., Haider, 1988) whereas the competing prenominal genitive is possible but stylistically marked (Fanselow, 2009). Accordingly, German is classified as a

language with dominant noun-genitive (i.e., postnominal) order, whereas English counts as a language without dominant order of genitive and noun (Dryer, 2013); thus, we are dealing with a language-specific subtype of syntactic inversion. Unlike the inversions in the Syntax condition, they are local in the sense that they occur within one of the major sentence constituents. Most of the canonical items in this condition conformed to the structure [determiner$_{NOM}$ – noun$_{NOM}$ – [determiner$_{GEN}$ – noun$_{GEN}$]], but we varied determiner types (e.g., articles, possessive determiners like *my*) and also included items with prenominal adjectives; see Table 4.3 for an example.

Table 4.3: Example couplets for the Genitive conditions. Boldface indicates lexical stress and/or phrasal accent; asterisks indicate optional stress.

(4a) Deviant: prenominal genitive

In [meiner lieben Eltern$_{GEN}$ Haus$_{NOM}$]$_{NP}$ war ich ein frohes Kind.
'*In my beloved parents? house, I was a cheerful child*.'

In	**mei**–	ner	**lie**–	ben	**El**–	tern	**Haus**
war*	**ich***	ein	**fro**–	hes	**Kind**		

(4b) Canonical: postnominal genitive

Im [Haus$_{NOM}$ meiner lieben Eltern$_{GEN}$]$_{NP}$ war ich ein frohes Kind.
'*In the house of my beloved parents, I was a cheerful child*.'

Im	**Haus**	mei–*	ner	**lie**–	ben	**El**–	tern
war*	**ich***	ein	**fro**–	hes	**Kind**		

GEN = genitive case; NOM = nominative case

Prenominal genitives make two major contributions to the metricality of the line: First, the noun-genitive inversion redistributes phonological words in a complex phrase. Second, prenominal genitives may reduce the number of unstressed syllables due to article omission (cf. English: '*the voice of the thunder*' vs. '*the thunder's voice*'). A subtler rhythmic effect is achieved by assigning phrasal

accent to the possessed instead of the possessor. Due to the locality of their effects, prenominal genitives typically correspond to line-internal metrical options that are frequently employed to avoid stress clashes or to align stresses and strong metrical positions as in the example in Table 4.3.

Agreement violations and unlicensed syntax (Type: Control)

The control condition reflected the conventional limits of poetic license and contained violations of grammatical constraints that usually cannot be dominated by metrical ones. The items in this condition were modifications of selected lines of verse and featured agreement violations or unlicensed syntactic inversions (see Table 4.4). Agreement violations were person, number, gender or case feature mismatches in subject–verb or modifier–noun agreement relations. Unlicensed inversions included illicit noun phrase discontinuities or inversions of determiner–noun order ('*the poem*' vs. '*poem the*') or adposition–noun order ('*in July*' vs. '*July in*'). Importantly, these deviations, too, affect the number and distribution of unstressed syllables and/or the arrangement of prosodic constituents. Exploiting these prosodic effects, we took care that our modifications satisfied metrical constraints just like the critical deviation types.

Participants and Procedure

We recruited 80 participants from the University of Frankfurt community (mean age = 24.55, SD = 5.68), all of them adult native speakers of German according to self-report (start of acquisition no later than age 3). Instruction groups matched in terms of age (two-sample t-test: $t(78) < 1$; $p > .90$) and represented a broad range of academic backgrounds; no further demographical data were collected. Participants received a campus-cafeteria coffee voucher as compensation. Each participant filled out a pen-and-paper rating questionnaire, which took about 10–15 minutes. Questionnaires contained a brief instruction and 5 familiarization items followed by 64 critical items presented in pseudo-randomized

Table 4.4: Example couplets for the Control conditions: agreement errors (4) and unlicensed syntactic inversions (5); differences are underlined. Boldface indicates lexical stress and/or phrasal accent; asterisks indicate optional stress.

(4a) Deviant: agreement error

Wieder, weil ein Jahr verging, schreiben man Sonette.
'*Again, because a year has passed, one are writing sonnets*.'

Wie–	der,	**weil***	ein	**Jahr**	ver–	**ging**,
schrei–	ben	**man**	So–	**net**–	te.

(4b) Canonical

Wieder, weil ein Jahr verging, schreibt man Sonette.
'*Again, because a year has passed, one is writing sonnets*.'

Wie–	der	**weil***	ein	**Jahr**	ver–	**ging**,
schreibt	man	So–	**net**–	te.

(5a) Deviant: illicit syntactic inversion

Ritter sitzt der liebestrunken zu den Füßen seiner Dame.
'*Knight is the sitting love-drunk at his lady's feet*.'

Rit–	ter	**sitzt**	der	**lie**–	bes–	**trun**–	ken
zu*	den	**Fü**–	ßen	**sei**–	ner	**Da**–	me.

(5b) Canonical

Der Ritter sitzt liebestrunken zu den Füßen seiner Dame.
'*The knight is sitting love-drunk at his lady's feet*.'

Der	**Rit**–	ter	sitzt	**lie**–	bes–	**trun**–	ken
zu*	den	**Fü**–	ßen	**sei**–	ner	**Da**–	me.

order. Participants were asked to read the sentences attentively and to judge either how *natural* (Task: acceptability, *n*=40) or how *poetic* they sound (Task: poeticity, *n*=40) by using the 7-point-scale next to each item; the scale ranged

from ‘not natural/poetic at all’ to ‘very natural/poetic’. All sentences, including couplets, were presented in a single line; this allowed us to isolate effects of rhythm and syntax from effects of lineation. Two further incomplete questionnaires were excluded from the analysis. After data collection, ratings were entered into a spreadsheet and double-checked for accuracy.

Data analysis

We tested for fixed main and interaction effects of Task, Type and Grammar on ratings using linear mixed-effects regression with crossed random effects for both participants and items (Baayen, Davidson & Bates, 2008). Significant interactions were resolved hierarchically; *p*-values of multiple comparisons were fdr-corrected; three-way interaction resolution involved fitting separate models for the two Task conditions. We used Satterthwaite’s method to approximate the degrees of freedom. Correlations (Pearson method, two-sided) between average acceptability and poeticity ratings estimated the relation between construction-specific deviance and poetic qualities for each deviation type. Analyses were carried out in *R* (R Core Team, 2016), using the packages *lme4* (Bates, Mächler, Bolker & Walker, 2015), *LMERConvenienceFunctions* (Tremblay & Ransijn, 2015) and *lmerTest* (Kuznetsova, Brockhoff & Christensen, 2017).

4.2.3 Results

Regression results

The overall model [Nagelkerke’s R^2: 0.49] revealed main effects of Type [$F(3,67.345) = 3.90$, $p = .013$] and Grammar [$F(1,81.917) = 8.80$, $p = .004$]. The former reflected differences between deviation classes, the latter a net penalty for deviant sentences [$t(81.9) = -2.97$]. Grammar differentially affected ratings in the two Tasks [significant interaction Task x Grammar: $F(1,89.264) = 113.02$, $p < .001$]: while acceptability dropped with the presence of deviant grammatical features [$t(88.6) = -9.51$, $p < .001$], poeticity increased [$t(83.0) = 6.03$, $p < .001$]. Net Deviation effects (across Tasks) differed between Types [significant interaction Type

x Grammar: $F(3,67.998) = 9.58$, $p < .001$]: in critical conditions, the antagonistic effects of Grammar on acceptability and poeticity cancelled each other out and thus neutralized any Grammar effect [all ps > .6]; the control condition, in contrast, showed a net disadvantage for deviant vs. canonical Grammar [$t(84.4) = -5.76$, $p < .001$].

Crucially, we observed an interaction of all three manipulated variables [significant interaction Task x Type x Grammar: $F(3,69.356) = 3.06$, $p = .034$]. Resolving the interaction by Task, we found that Grammar effects on ratings differed across Type conditions in both the acceptability Task [$F(3,2372.09) = 37.94$, $p < .001$] and the poeticity Task [$F(3,81.659) = 3.57$, $p = .018$]. Mean ratings and results of planned comparisons are provided in Table 4.5. Deviant sentences were rated less acceptable than non-deviant ones across all Type conditions; deviant controls showed a particularly heavy penalty. Poeticity ratings, in contrast, were higher for deviant sentences than for non-deviant ones in all three critical Type conditions; the control condition showed no effect of Grammar on poeticity (see Figure 4.1).

Table 4.5: Mean ratings (±SEM) per condition and results of planned comparisons; rating scales ranged from 1 (not acceptable/poetic at all) to 7 (very acceptable/poetic).

	Acceptability				Poeticity			
	Deviant	Canonical	*t*		Deviant	Canonical	*t*	
Control	2.71 (±0.16)	4.76 (±0.16)	–14.86	***	4.15 (±0.14)	3.86 (±0.18)	1.55	
Syntax	3.81 (±0.14)	4.62 (±0.19)	–5.96	***	4.25 (±0.12)	3.34 (±0.13)	7.51	***
Morphology	3.60 (±0.12)	4.31 (±0.13)	–5.24	***	4.95 (±0.13)	4.28 (±0.14)	5.52	***
Genitive	3.73 (±0.21)	4.31 (±0.15)	–4.34	**	4.86 (±0.15)	4.27 (±0.19)	4.39	***

* <.05; ** <.01; *** <.001

Correlation results

We tested for correlations between acceptability and poeticity ratings in all Type conditions. As illustrated in Figure 4.2, there was no consistent association in the control condition, but all critical conditions showed negative correlations be-

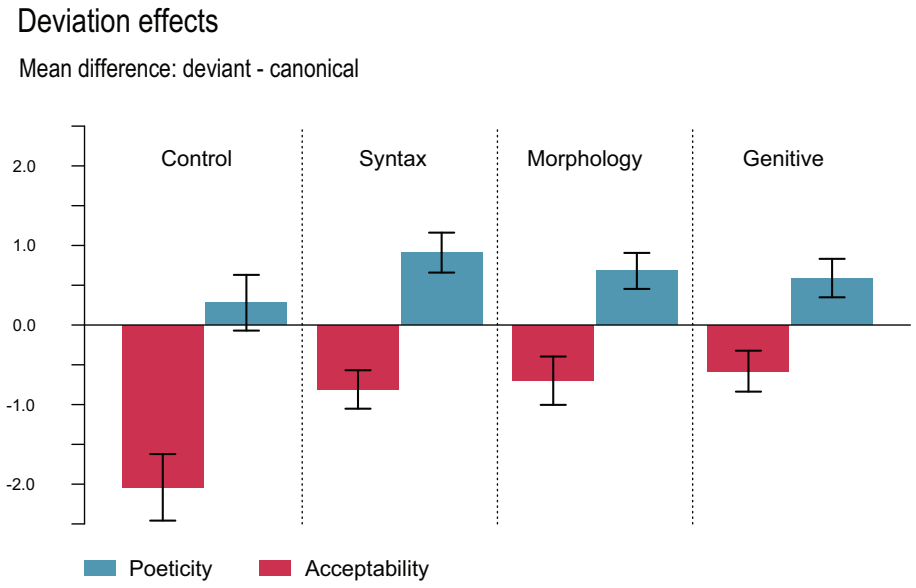

Figure 4.1: Results of Experiment 3a. Mean deviation effects on ratings by Task (color-coded) and deviation Type (panels). Bars indicate mean differences (deviant – canonical versions); error bars indicate 95% confidence intervals.

tween these two qualities, i.e., less acceptable sentences were perceived as more poetic and vice versa. We found a strong correlation effect for syntactic inversions and effects of moderate strength for both the morphological deviation class and the prenominal genitives; see Table 4.6 for details.

Table 4.6: Results of Experiment 3a. Correlation strengths (and 95%-CIs) by deviation type.

	df	CI_{lower}	*r*	CI_{upper}	p_{fdr}	
Control	30	–.43	–.11	.25	.561	
Syntax	30	–.78	–.60	–.31	.001	**
Morphology	30	–.68	–.43	–.09	.019	*
Genitive	30	–.69	–.45	–.11	.019	*

* <.05; ** <.01

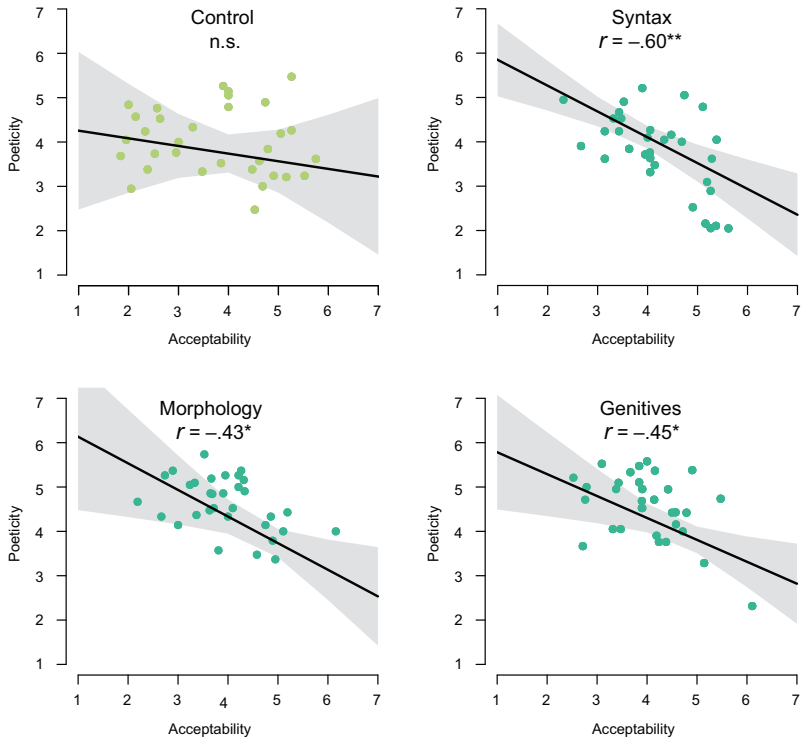

Figure 4.2: Results of Exp. 3a. Correlations of poeticity and acceptability ratings for the four deviation types. Shades indicate the 95% confidence band.

4.2.4 Discussion of Experiment 3a

We compared effects of poetically licensed (=critical) and unlicensed (=control) deviations on the acceptability and poeticity of selected lines of verse; both critical and control deviations resulted in parallelistic sentence rhythm. The presence of licensed grammatical deviations decreased the acceptability of the lines while increasing their poeticity; these conditions also revealed an association between deviance (i.e., low acceptability) and greater poeticity. Control deviations patterned differently in both tasks and showed a particularly severe acceptability penalty but no effect on the poeticity of the lines; deviance and poeticity were unrelated in this condition. This pattern of results corroborates our distinction between licensed and unlicensed deviations and suggests categorical dif-

ferences between these conditions in terms of grammatical and genre-specific well-formedness. It replicates Martindale's (1973) finding of an upper limit to the poeticity of grammatical deviance but, contrary to Martindale, we observed that this limit coincides with the boundary of grammatical well-formedness – at least without the stylistic context of an entire poem. The observed pattern further supports the view that contemporary readers draw on genre-specific knowledge when judging the poeticity of unfamiliar utterances, and that they favor deviations that bend the rules of grammar over those that break them.

The considerable drop in acceptability observed for control deviations indexes the categorical ungrammaticality of agreement errors and unlicensed syntactic inversions, i.e., their non-conformity with the strict rules of grammar. Considering that metrical benefits were comparable for licensed and unlicensed deviations, the absence of a poeticity effect in the control condition most likely reflects their non-conformity with genre conventions. They could, of course, occur in poetry, but they are not prototypically associated with the genre, and thus not licensed a priori. The extremely low acceptability of these items reflects their high disruptive potential; the absence of a poetic license suggests that the effect on the reader will be one of severe deautomatization if a poet deliberately employs such deviations for stylistic effect – the poetry of E. E. Cummings provides some well-known effective examples.

The gradient decrease in acceptability observed for licensed deviations reflects degrees of grammatical well-formedness: while virtually all critical sentence versions conform to the rules of grammar, deviant alternatives are perceived as more marked, unusual and striking than canonical ones, and their reduced acceptability suggests increased processing difficulty during reading. Analogously, the gradient poeticity differences seem to index degrees of genre-specific well-formedness: while all critical versions conform to the conventions of the genre, deviant alternatives are perceived as more poetic than their canonical counterparts, which we interpret as evidence that these deviations sound more prototypically poem-like, and that their effects on online comprehension contribute to a distinctly poetic processing experience (Castiglione, 2017; Hoffstaedter, 1987).

The genre prototypicality component relates to previous work by Hanauer (1995, 1996), who identified classes of cues that influence readers' text categorization judgments. He demonstrated that literary expertise affects the weightings of cues from different strata of text organization (e.g., graphic, phonological, syntactic, thematic). Our experiment examined whether and to which degree specifically deviant features from several strata of sentential organization (i.e., word, phrase and sentence) are associated with the schematic genre prototype acquired by experienced readers. Sampling from a dominantly non-expert student population, we found that combined syntactic and prosodic cues strongly affected sentential poeticity ratings. At first glance, this finding may seem inconsistent with Hanauer's observation that self-reported reliance on syntactic poeticity cues is quite low across levels of expertise. However, we assume that this asymmetry merely reflects resolution and cue availability, i.e., differences in the properties of the materials (texts vs. sentences). While effects of literary expertise were beyond the scope of the present investigation, our experience-based conception of genre concedes what Hanauer's research suggests, namely that expertise is a key variable in literary-aesthetic evaluation whose influence deserves to be systematically addressed in future research. The processing component of the poeticity judgment relates to Hoffstaedter's (1987) investigation of "poetic text processing". Arguing for a processing-based conception of poeticity, Hoffstaedter attributed effects of parallelism and elliptical constructions on text-level poeticity to subevents during reading. In this view, our effects reflect the processing of sentential syntax and (implicit) sentence rhythms; this point will be elaborated below.

The combined effects of parallelistic meter and deviant grammar support foregrounding theories of literary language, which trace distinctly literary reading experiences to the deautomatization induced by deviation and parallelism (e.g., Miall & Kuiken, 1994; Mukařovský, 1964; van Peer, 2007). Notably, the results in our critical conditions provide empirical support for Mukařovský's (1964:21ff) claim that linguistic features and constructions may be foregrounded in the written standard (as indexed by acceptability) but "unforegrounded" in the literary canon (as indexed by poeticity). From such a pattern, we may derive the hy-

pothesis that the disruptive potential of routine licenses will not fully develop in a poem; the adverse cognitive effects of some licensed morphological deviations, at least, are neutralized if readers are aware of the text's genre (Blohm, Menninghaus & Schlesewsky, 2017).

Note that these results are not straightforwardly generalizable to other languages and literary traditions. We would expect to find a similar pattern for deviations in English verse, though, due to the relatedness of the languages and literatures. Still, deautomatization of grammatical processing depends both on the grammar of a language and on readers' comprehension systems. For instance, English employs word order to encode basic relations between predicates and arguments, whereas German relies primarily on case marking and agreement. Consequently, English native speakers rely strongly on syntactic cues during online sentence comprehension ("who did what to whom?") whereas German native speakers are particularly sensitive to morphology (e.g., Bornkessel-Schlesewsky & Schlesewsky, 2009; MacWhinney, Bates & Kliegl, 1984). In this sense, the cognitive and aesthetic effects of grammatical deviation are language-specific: native speakers of English and German could well differ in their sensitivity to syntactic and morphological deviations.

Finally, the observed pattern of results resembles recent evidence regarding effects of rhetorical features on readers' aesthetic evaluation of verbal stimuli. Examining effects of rhyme and brevity/meter on the aesthetic evaluation of proverbs, Menninghaus and colleagues (2015) found that these rhetorical features reduce comprehensibility while increasing the perceived beauty and persuasiveness of the proverbs. They interpreted their findings in the framework of fluency-based aesthetics (e.g., Alter, 2013; Bullot & Reber, 2013; Graf & Landwehr, 2015), which links aesthetic liking to both cognitive fluency and disfluency at distinct stages of stimulus processing; they concluded that combined perceptual processing fluency and conceptual disfluency are both pleasing and effective. Our results align with these preferences: easy-to-process rhythms and hard-to-process grammar increase the poeticity of isolated lines or couplets, suggesting that this particular combination is central to the prototypical experience of reading poetry. Following this conception and considering that meter

has aesthetically appealing qualities of its own (e.g., Menninghaus et al., 2017; Obermeier et al., 2013; van Peer, 1990), it seems desirable to dissociate the poetic effects of fluency-reducing deviation and fluency-enhancing parallelism.

4.3 Experiment 3b: Syntax and prosody in regular sentences

Experiment 3b focused on naturally occurring grammatical variation in regular sentences, i.e., on "the creative aspect of poetic language" (Leech, 1969:23). It supplements the syntax condition of Experiment 3a while avoiding the clear lexical, thematic, and grammatical genre cues that were present in the materials of the previous experiment. We reasoned that this difference would reduce the genre-prototypicality component of the poeticity judgment and increase the qualitative component that reflects the evaluation of well-composed processing trajectories. We used single-item two-alternative forced choice (2AFC) to test (a) whether the characteristic combination of perceptual fluency and conceptual disfluency leads readers to assign poetic qualities to syntactic configurations that occur naturally as part of German grammar, and (b) whether the effects of parallelistic rhythm and non-canonical grammar could be dissociated.

4.3.1 Design and hypotheses

As in Experiment 3a, we manipulated Task (acceptability vs. poeticity) between participants. We constructed four variants of a single sentence, crossing Syntax (canonical vs. non-canonical) and sentence Rhythm (alternating vs. non-alternating). The single-item design does not allow for statistical generalization across sentences. However, since the four variants did not differ regarding word choice or truth-conditional meaning, we maintain that differences between conditions are most likely due to syntactic/prosodic factors. Based on the results of Experiment , and on the findings of Menninghaus and colleagues (2015), we expected greater poetic quality for (a) strictly alternating (vs. non-alternating) rhythm patterns that facilitate prosodic processing (e.g., Bohn, Knaus, Wiese &

Domahs, 2013; Kriukova & Mani, 2016), and for (b) non-canonical (vs. canonical) word orders that complicate syntactic processing and reduce both acceptability and ease of comprehension (e.g., Pechmann, Uszkoreit, Engelkamp & Zerbst, 1994; Rösler et al., 1998).

4.3.2 Materials and methods

Materials

In German declarative main clauses, any single major sentence constituent may occur in the prefield, i.e., in the clause-initial syntactic position preceding the finite verb (Haider, 2010:1). In everyday communication, discourse requirements govern which element actually occupies the prefield (e.g., Speyer, 2004). Out of context, however, frame-setting sentence adverbials and prototypical sentence topics, e.g., pronominal grammatical subjects, are preferred in this position. In each of the four versions (see Table 4.7), we fronted one of the non-verbal constituents to the prefield while leaving the other constituents in their canonical post-verbal positions.

Sentences (1) and (2) are syntactically canonical in the sense that they conform to the canonical linearization of the nominal predicate 'a riddle' and its pronominal arguments 'you' and 'to-me' (Syntax: canonical). Sentences (3) and (4) disobey violable grammatical (nominative before dative case) and pragmatic (given before new information) linearization principles (Syntax: non-canonical; cf. Uszkoreit, 1986). In sentences (2) and (4), stressed and unstressed syllables alternate regularly (Rhythm: alternating), whereas sentences (1) and (3) feature a non-alternating rhythm pattern with two stress lapses (Rhythm: non-alternating). From this quadruple (1–4), we created all 12 possible permutations of ordered sentence pairs.

Participants and Procedure

We recruited 240 self-reported adult native speakers of German (start of acquisition no later than age 3) on the campus of Goethe University Frankfurt,

Table 4.7: Stimulus sentences in Experiment 3b. Boldface highlights prominent syllables.

(1) canonical$_{\text{Syntax}}$ – non-alternating$_{\text{Rhythm}}$

Du	**bist**	mir	in	**man**–	chen	**Ding**–	en	ein	**Rä**–	tsel.
you	are	to-me	in	some		things		a	riddle	

You are a riddle to me in some respects.

(2) canonical$_{\text{Syntax}}$ – alternating$_{\text{Rhythm}}$

In	**man**–	chen	**DIng**–	en	**bist**	du	**mir**	ein	**Rä**–	tsel.
in	some		things		are	you	to-me	a	riddle	

In some respects, you are a riddle to me.

(3) non-canonical$_{\text{Syntax}}$ – non-alternating$_{\text{Rhythm}}$

Mir	**bist**	du	in	**man**–	chen	**DIng**–	en	ein	**Rä**–	tsel.
to-me	are	you	in	some		things		a	riddle	

To me, you are a riddle in some respects.

(4) non-canonical$_{\text{Syntax}}$ – alternating$_{\text{Rhythm}}$

Ein	**Rä**–	tsel	**bist**	du	**mir**	in	**man**–	chen	**DIng**–	en.
a	riddle		are	you	to-me	in	some		things	

A riddle you are to me in some respects.

Germany; no demographical data were collected. Each participant received a sheet of paper with one pair of sentences. Participants were instructed to read the sentences attentively and to mark intuitively the more natural-sounding one of the pair (Task: acceptability, *n*=120; cf. Clifton, Fanselow & Frazier, 2006; Haupt, Schlesewsky, Roehm, Friederici & Bornkessel-Schlesewsky, 2008) or the more poetic-sounding one (Task: poeticity, *n*=120). Ten participants per Task condition judged each ordered pair.

Data analysis

For each ordered pair, we coded whether the Syntax of the two sentences was canonical^{1st}–non-canonical^{2nd} (+1), non-canonical^{1st}–canonical^{2nd} (–1), or equal in terms of canonicity (0), and whether their Rhythm was alternating^{1st}–non-alternating^{2nd} (+1), non-alternating^{1st}–alternating^{2nd} (–1) or equal (0). Task was contrast-coded (acceptability = –1; poeticity = 1). We analyzed choices (1st or 2nd sentence of the pair) using logistic regression with Syntax, Rhythm, Task and their interactions as predictors. Significant interactions were resolved hierarchically by separate analyses for the Task conditions. Analyses were carried out in *R* (R Core Team, 2016).

4.3.3 Results

The observed choices indicate that the tasks differ in their word order preferences: acceptability choices clearly favored sentences with canonical Syntax (i.e., 1 & 2), whereas poeticity choices showed a less pronounced preference for non-canonical Syntax (i.e., 3 & 4); see Table 4.8. The overall model [Nagelkerke R^2 = .65] confirmed the expected interaction of Task x Syntax [z = –7.04, p < .001, OR (odds ratio) = 0.05, $[0.02, 0.10]_{95\%CI}$]. As shown in Figure 4.3A, there was a strong preference for canonical word orders in the acceptability task [z = 5.16, p < .001, OR = 42.54, $[13.11, 262.58]_{95\%CI}$], whereas non-canonical syntax was favored in the poeticity task [z = –5.06, p < .001, OR = 0.11, $[0.04, 0.23]_{95\%CI}$].

A main effect of Rhythm [z = 3.35, p < .001, OR = 2.38 $[1.48, 4.20]_{95\%CI}$] was qualified by an interaction with sentence Rhythm [z = 2.98, p = .003, OR = 2.17 $[1.34, 3.81]_{95\%CI}$]. As illustrated in Figure 4.3B, there was no effect of Rhythm on acceptability [z = 0.31, p = .756, OR = 1.10 $[0.60, 2.04]_{95\%CI}$], but poeticity choices favored alternating over non-alternating Rhythm [z = 3.95, p < .001, OR = 5.16 $[2.51, 13.73]_{95\%CI}$].

Table 4.8: Results of Experiment 3b. Number and percentage of choices by condition and task.

	Condition		Acceptability		Poeticity	
	Syntax	Rhythm	*n*	%	*n*	%
(1)	canonical	non-alternating	44	73.3	8	13.3
(2)	canonical	alternating	54	90.0	26	43.3
(3)	non-canonical	non-alternating	15	25.0	35	58.3
(4)	non-canonical	alternating	7	11.7	51	85.0
		Relative ranking:	(2) > (1) > (3) > (4)		(4) > (3) > (2) > (1)	

4.3.4 Discussion of Experiment 3b

Reflecting only syntactic canonicity, the acceptability results replicate the well-known penalty for non-canonical word orders (for German e.g., Pechmann et al, 1994). By contrast, poeticity choices were sensitive to both syntax and rhythm, and favored non-canonical (i.e., less acceptable) syntactic structures and/or alternating rhythms. Note that the canonical variants (1) and (2) show the same relative ranking in both tasks; this pattern refutes a potential interpretation of the poetic rhythm effect as a gradient effect of syntax, which would incorrectly predict (1) to score higher on poeticity than (2). Above all, these results confirm that readers have consistent intuitions about the poetic qualities of regular sentences; they replicate the effects of word order and patterned rhythm on the poeticity of verse lines and extend them to syntactic and prosodic configurations that occur naturally as part of contemporary German grammar. Taking into account syntactic variation and subtle prosodic differences, readers seem to apply evaluation criteria consistently when judging the poetic quality of verbal stimuli irrespective of the presence of clear genre markers. Notably, our findings dissociate the effects of parallelism and deviation and reveal that syntactic canonicity has greater impact on poeticity choices than implicit prosody. This possibly reflects that syntactic variation is cognitively salient and more accessible to post-hoc evaluation than the subtle prosodic variations of readers' inner voices. Alternatively, this finding may indicate that the deautomatization of form-to-meaning mapping is more central to contemporary readers' conception

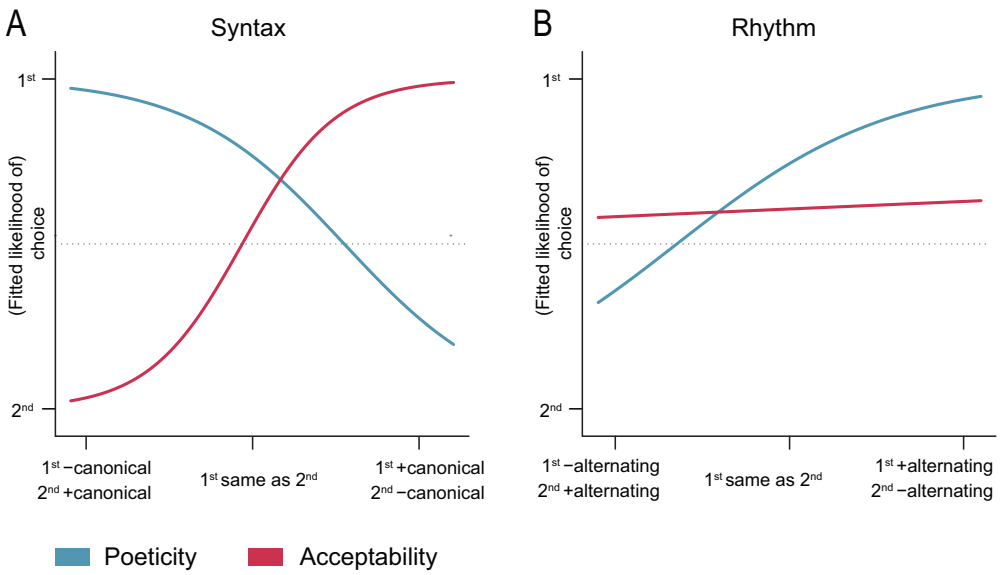

Figure 4.3: Results of Experiment 3b. Fitted choice likelihood for the two tasks by Syntax (A) and Rhythm (B). The dotted horizontal line reflects chance level; solid lines indicate how choices vary with the presence/absence of syntactic and prosodic cues.

of poeticity than the regularity of meter, at least at the sentence/verse level. In any case, the results of Experiment 3b further corroborate the idea that both facilitated prosodic processing and increased cognitive effort during form-to-meaning mapping characterize a distinctly poetic reading experience.

4.4 Discussion and conclusion

In two experiments, participants read couplets of selected verse (Experiment 3a) or regular sentences (Experiment 3b) and judged them for either acceptability or poeticity. To our knowledge, these studies are the first to combine grammatical and literary-aesthetic intuitions in a single experimental paradigm; their focus on deviant sentential grammar complements related work that used sentence judgment methods to investigate rhetorical and aesthetic effects of parallelism (e.g., McGlone & Tofighbakhsh, 1999; Menninghaus et al., 2015). Our experiments

converge in the finding that prosodic parallelism and non-canonical grammar (but not utter ungrammaticality) increase the perceived poeticity of verbal stimuli. Our second experiment provides initial evidence that these effects indeed reflect dissociable contributions to poeticity, which follows naturally from an account that relates poetic effects of grammar and rhythm to underlying processing events at distinct stages of comprehension.

Poeticity judgments were sensitive to grammatical features and to subtle variations in implicit prosody; they showed comparable judgment patterns in response to actual verse and to regular sentences. In contrast, acceptability reflected only grammatical variation and penalized non-canonicity and ungrammaticality. To us, this pattern suggests that the poeticity task (but not the acceptability task) triggers a reading stance that comes with its own set of grammatical norms (Mukařovský, 1964) and that possibly involves increased attention to sound recurrences (Beaugrande, 1978; Rosenblatt, 1978, 1988).

A serious objection could be raised regarding the appropriateness of sentence judgments for addressing questions of poetic language, namely that measuring the effects in isolated verse lines and sentences neglects that poems are texts with notoriously intricate phonological, grammatical and semantic relations between lines. However, we maintain that establishing the context-free deviance and poeticity of grammatical constructions should precede the investigation of how they vary as a function of context. Besides, the rough equivalence of clauses/sentences, i.e., central units of grammatical structure, and verse lines, i.e., central units of poetic structure, theoretically supports the linguistic study of how poetic forms constrain grammatical form (e.g., Youmans, 1983).

Another concern relates to the level of poetic competence which, compared to linguistic competence, emerges from a relatively sparse input and is thus subject to greater inter-individual variation. While literary expertise is doubtlessly a key determinant of the literary-aesthetic evaluation of verbal stimuli, it was not the object of investigation here. We used random sampling techniques to warrant the assumption that the average level of expertise in our samples approximated the mean of the population of educated (=student) readers. This

is presumably a basic level (Hanauer, 1995), and it remains unclear in how far sentence evaluation varies with the degree of expertise; this is left to suitably designed future studies.

Still, what can poeticity judgments reveal about poetry if they merely reflect the absence/presence of pre-selected genre-typical linguistic features? First, they potentially capture stylistic effects of completely novel forms and features. As novel forms and techniques are appreciated and to a certain extent expected in poetry, we assume that they are reflected in these judgments if they are stylistically effective. Our second experiment showed, at least, that known forms, i.e., syntactic inversions, continue to be effective in novel combinations and contemporary contexts. More importantly, sentence judgments partially reflect properties of the act of reading, i.e., they inform us about reader response to text features not about the features themselves. After all, what is accessible to readers, and what constitutes the object of their judgment, is not the "text itself" but the experience of reading and the resulting (and intermediate) mental representations. Syntactic inversions, for instance, affect the linear order of meaningful elements, and thus the temporal order in which conceptual nodes are activated in readers' minds. This determines the (degree of) engagement of cognitive processes necessary to construct a coherent representation of the message conveyed. Deviations from canonical word order necessitate departures from automatized processing routines (e.g. Cureton, 1980; Dillon, 1978; Rösler et al., 1999). Acceptability judgments are sensitive to such cognitive events during comprehension, e.g., misanalyses or ambiguity (e.g., Fanselow & Frisch, 2006; Hofmeister, Jaeger, Arnon, Sag & Snider, 2013), and we assume that poeticity judgments reflect them, too (Hoffstaedter, 1987). In fact, our data seem to suggest that it is the very same processing events that reduce an utterance's acceptability and increase its poeticity. Although only indirectly derivable from our results, we further assume that these judgments reflect stylistically effective processing trajectories. As an illustration, consider sentence (4) from Experiment 3b, which was judged the least acceptable and the most poetic of the four sentence versions:

Ein Rätsel bist du mir in manchen Dingen.
lit: a riddle are you to-me in some things
'*A riddle you are to me in some respects*.'

The rhythm of this sentence is alternating and easy to process, allowing readers to predict prosodic features of incoming material. While prosodic processing runs fluently, conceptual processing is disrupted early on: the sentence-initial constituent '*a riddle*' is formally compatible with the grammatical subject function and is preferably interpreted as such (e.g., Hemforth, 1993). In the absence of evidence to the contrary, the processing system will, at this point, make the minimal assumption that '*a riddle*' is the sole argument of a one-place predicate, e.g., an intransitive verb or an adjective as in '*A riddle can be solved/is solvable*'. Immediately upon encountering the copula verb '*bist*' (= 'are_{2SG}'), however, the agreement marking (2nd person singular) renders the initial interpretation untenable and requires (a) reanalyzing the assumed grammatical subject as a nominal predicate, and (b) predicting an as yet unknown argument of this predicate. Reanalysis is straightforward, since the unambiguous subject '*du*' (= '*you*') follows immediately. This mis- and reanalysis pattern is absent in the non-canonical sentence (3), and it might partially account for both the relative acceptability and poeticity of (3) and (4).

However, reading sentence (4) involves more than mis- and reanalysis: a particularly poetic processing experience may result if the mental operations performed resemble the content of the representations involved, forcing the reader to re-enact the meaning of the text (Fish, 1970). The concept 'riddle' implies solvability, i.e., the possible transition from non-understanding to understanding. In (4), this concept has just been activated in the mind of the reader when he is detecting and resolving an error in his initial syntactic analysis. The detected error is not merely a grammatical one, though: along with the re-interpretation of the grammatical subject as the predicate of the sentence, reanalysis involves coercing the literal meaning of '*a riddle*' into the metaphorical meaning of the idiomatic expression which denotes a state or relation of non-understanding between a non-understander and a non-understood. Following this initial inter-

pretive tension-release pattern, integration of the rest of the sentence is fairly easy again. The case marking on the pronoun '*mir*' (= '*to-me*') indicates unambiguously that it refers to the non-understander, which is the role the silently articulating reader associates with if he adopts the speaker's perspective. The final phrase '*in manchen Dingen*' (= '*in some respects*') then restricts the statement blandly. The analyses, predictions, error detections and conflict resolutions, the convergence of operations and representation in the transition of non-understanding to understanding – all of this happens in a few seconds. Potentially, all of this also co-determines a reader's poeticity judgment, which opens the intriguing prospect that these judgments not only reflect prototypically poetic configurations that combine perceptual fluency with conceptual disfluency, but also more nuanced processing trajectories that fuse meaning and experience. If this assumption proves to be correct, experimental syntax and experimental poetics are facing, for different reasons, the same need to dissociate competence from performance. Whereas processing effects interfere with the attempt to isolate the grammatical rule system in experimental syntax, the rule-like system of genre conventions interferes with the attempt to isolating stylistically effective processing events in experimental poetics.

We used sentence judgments to study how parallelistic and deviant linguistic features affect readers' grammatical and literary-aesthetic evaluation of verse lines and regular sentences. In terms of theory, the results support the view that parallelism-induced perceptual fluency and deviation-induced conceptual disfluency are prototypical ingredients of the experience of reading poetry, and that the latter may even be more central than the former. In terms of method, they illustrate how the parallel collection of readers' linguistic and literary-aesthetic intuitions provides fine-grained objective measures that inform about the relation between linguistic structure and poetic effect. Sentence judgments allow studying this relation systematically and in more detail than currently practiced in experimental poetics. Possible future applications of these methods include (a) a multi-item validation of our single-item results to corroborate the initial evidence for dissociable poetic effects of implicit sentence rhythm and non-canonical

grammar, (b) an investigation of whether poeticity judgments are indeed sensitive to ambiguity and reanalysis, which would further support our assumption that these judgments indeed reflect properties of sentence processing, (c) investigations of how literary expertise affects the evaluation of verbal stimuli, (d) the dissociation of poetic effects of deviation and parallelism in actual verse, (e) examinations of the relation between linguistic structure and other, more clearly aesthetic evaluative dimensions such as beauty, liking, or melodiousness.

In conclusion, we propose that the systematic collection of readers' intuitive judgments constitutes a step towards an experimental "affective stylistics" (Fish, 1970), regardless of the specific question addressed. The combined collection of linguistic intuitions and poeticity judgments, as illustrated here, helps sharpening our understanding of how 'ordinary' language use relates to poetic language use.

5
Conclusion

Dedicated to the study of literary non-experts' conceptions of the poem, this thesis presented a series of experiments that used behavioral and electrophysiological measures to shed light on different aspects and stages of the literary reception process. We will briefly recapitulate the major findings, not in their previous order but—relating them to distinct stages of literary comprehension––in the order in which they occur during reading.

The **immediate pre-receptive phase**, i.e., the phase when the verbal stimulus is expected but not yet perceived and processed, was the focus of *Experiment 1c*, in which we analyzed oscillatory activity in the alpha band to investigate whether genre categorization modulates anticipatory attentional states. We observed greater induced pre-stimulus alpha power in the poetry group than in a neutrally instructed control group. This finding aligns with effects of cued selective attention in other domains, including tactile and visual perception, and it seems to indicate that genre categorization triggers the functional inhibition of cortical areas that are involved in sentence processing under neutral conditions but that are irrelevant for the task of reading poetry. This anticipatory gating of information thus lends empirical support to the notion that text categorization

triggers genre-specific expectations which induce selective attention focused on specific—but as yet unknown—aspects of the written verbal input.

The **initial reading phase** was studied in *Experiments 1b/1c*, which used event-related brain responses (ERPs) to investigate *a priori* effects of genre categorization on online sentence comprehension. We observed greater N300 amplitudes in response to the sentence-initial constituent in the poetry group than in the neutrally instructed control group. While the exact functional significance of this effect remains elusive, we took its insensitivity to our morpho-phonological and semantic manipulations as well as its correlation with pre-stimulus alpha power as indicators that it relates to selective attention which—we suggested—focuses on the extraction of phonological information from written words during reading. While this functional interpretation is, so far, only supported by related ERP evidence and the results of *Experiment 3b*, the occurrence of this effect demonstrates that genre-induced differences emerge very early after the onset of text presentation.

Historical vs. contemporary case suffixes in the sentence-initial constituent modulated word prosody and thus the regularity of implicit sentence rhythm. Historical inflections decreased the amplitude of a left-lateralized N400-like negativity in response to the following word. Based on related results that show prosodic effects during silent sentence reading, we interpreted this reduction as a facilitative effect of prosodic regularity on word recognition; its insensitivity to genre categorization fails to support the notion that reading poetry entails increased attention to sound recurrences independent of their actual occurrence, an idea that was further investigated in *Experiment 2*. Although archaic case suffixes facilitated lexical processing of following words, they decreased accuracy during concurrent task performance, at least when readers were unaware of the genre. These results indicate that deviant morphological features may impair and facilitate distinct aspects and stages of stimulus processing during reading.

Finally, semantically incongruent vs. congruent subject noun phrases elicited an N400 effect followed by a late positivity whose occurrence was delayed in

poetry readers. This pattern suggests that early semantic processing is unaffected by genre categorization, but that late pragmatic-evaluative processes are subject to genre-specific top-down control; this interpretation is further corroborated by the meaningfulness judgments collected in *Experiment 1a*, which indicated that the genre of poetry entails increased interpretive effort in the face of apparent semantic incongruity. The idea that genre categorization modulates prosodically and semantically mediated lexical access was not supported by the results.

The **reading phase** was further examined in *Experiment 2*, which used eye tracking and speech recordings to investigate how genre categorization affects reading behavior and text performance during oral reading. In line with our expectations we observed a range of systematic differences between poetry and literary prose in the eye-movement data and in the speech signals. Since these differences were independent of linguistic variables, they provide strong evidence for the conceptual differentiation of these basic literary categories and for the behavioral differentiation of the respective material-appropriate reading strategies. Contrary to previous studies, this experiment not only sketched behavioral and phonetic profiles associated with these literary genres but also focused on distinctive dynamics of genre-appropriate reading. An increasing tendency to skip short words was observed during prose reading, reflecting that accumulating discourse-contextual constraint usually warrants less careful navigation through the text; this tendency was greatly attenuated in poetry reading—particularly line-initially but also line-internally—which suggests that the reading strategy for poetry is less risky and does not fully exploit the available discourse-contextual information, and also that the graphic conventions of the genre support more careful reading.

Focusing on the early interaction of genre categorization and prosodic context, we obtained no support for the idea that reading poetry involves increased attention to prosodic sound recurrences; however, we observed that poetry readers emphasize linguistic rhythm by selective word lengthening. While the metrical importance of monosyllabic words has been pointed out repeatedly (e.g.,

Kiparsky, 1975; Fabb, 2001), our empirical results identified monosyllabic function words as a lexical class whose prosodic properties are exploited even by literary non-experts for the imposition of metrical accent.

The **immediate post-receptive phase** was the focus of *Experiment 1a* and *Experiments 3a/3b*, which used systematically collected sentence judgments to measure readers' subjective impressions immediately after reading but with enough time for conscious deliberation. *Experiment 1a* revealed that poetry categorization increases the perceived meaningfulness of verbal stimuli. It was particularly the utterances which seem semantically incongruent at first glance that appeared more meaningful to readers in a poetic context than in a non-literary one. This result and the corresponding late positivity effect observed in the ERP study align with earlier findings (Gibbs et al., 1991; Zwaan, 1993) and support the *polyvalence convention* hypothesis (e.g. Meutsch & Schmidt, 1985), which maintains that literary comprehenders are free to explore and pursue more than one interpretation, including idiosyncratic ones.

Experiment 3a compared the evaluative stance of the contemporary written standard to the evaluative stance of poetry. Analyses of systematically collected acceptability and poeticity judgments for selected and modified lines of verse revealed that different classes of conventionally licensed grammatical deviations add to varying degrees to the perceived poeticity of a sentence, but that unlicensed ungrammaticalities have no such effect. We argued that intuitive evaluations of poeticity reflect how poem-like a sentence sounds, but that they also capture a perceived quality of the reading experience which is grounded in the processing events triggered by licensed deviations and metrical regularity. In any case, the observed pattern suggests that only selected deviant grammatical constructions and features are associated with readers' schematic conceptions of the poem. Poetry readers thus bring knowledge of expectable and permissible grammatical structures to the act of reading, as well as increased but limited tolerance for grammatical deviation; contrary to prior findings in English, this limit coincided with unambiguous ungrammaticality, e.g., agreement errors. Furthermore, we observed in *Experiment 1b* that grammatical but marked morpho-

phonological features impaired concurrent task performance during sentence reading, but that genre categorization neutralized this effect. These findings indicate that poetry readers do not merely tolerate more grammatical oddities. It rather seems that they are more *resilient* and in fact prepared—due to pre-activated genre knowledge—to process licensed grammatical deviations without suffering the interference observed in sentence reading under neutral conditions, which supports the notion of foregrounding as register-inappropriate verbal behaviour (Havránek, 1964) and the idea that foregrounding in poetry depends on the norms of both the standard language and the poetic canon or genre (Mukařovský, 1964).

Finally, *Experiment 3b* used single-item 2AFC to provide initial evidence that readers employ comparable evaluation criteria when judging the poetic qualities of regular sentences that lack clear linguistic genre cues. Results indicated that similar criteria were applied to linguistic variation that occurs naturally as part of contemporary German grammar. Readers took not only the grammatical markedness of the sentences into account, but also the implicit sentence rhythm: marked grammatical structures that render form-to-meaning mapping more challenging increased poeticity as did sentence rhythms whose expectation-inducing regularity facilitates prosodic processing during reading. This is, in fact, the only finding of the presented experiments that lends support to the idea that reading poetry involves increased attention to sound recurrences; the absence of supporting evidence in the more challenging tasks, i.e., rapid sentence reading and probe recognition (*Exp. 1b*) or oral reading (*Exp. 2*), suggests that literary non-experts require conscious evaluation (*Exp. 3b*) to recognize subtle variations in implicit sentence prosody.

These observed neural and behavioral adjustments demonstrate that readers' genre conceptions influence all stages of the reception process: anticipatory attentional states in the immediate pre-receptive phase, processing strategies in the actual reading phase, and evaluation criteria for verbal stimuli in the immediate post-receptive phase. We found no support for the widely held assumption that reading poetry entails increased attention to sound recurrences in the

assumption that these are not random—at least not within the investigated student population and with respect to the prosodic recurrences of meter. However, when these readers were given enough time to reflect on the verbal stimulus, and when they were asked to consciously evaluate its poetic qualities, then they took subtle prosodic variation in implicit sentence rhythms into account. In other words: the subtleties of verse may be noted by literary non-experts only in situations that more closely resemble a reading-for-leisure situation, where the reader is in control—not poets, teachers, or experimenters.

Based on our findings, we can confirm the following characteristics of the reading strategy for poetry as triggered by genre categorization:

1. **Adjusted expectations and focal attention prior to reading** (*Exp. 1c*), but no support for increased attention to prosodic recurrences in reading (*Exp. 1b*, *Exp. 2*)
2. **Adjusted information gathering routines and eye-movement control** (*Exp. 2*)
3. **Genre-specific phonetic recoding** of phonological structure (*Exp. 2*); the poetry schema imposes an additional filter on phonetic output
4. **Resilience to conventionally licensed grammatical deviation** (*Exp. 3a*, behavioral task in *Exp. 1b*); ungrammaticality is not licensed
5. **Altered standards of evaluating linguistic form and semantic congruence** (*Exps. 1a, 3a/b*):
   Grammatical markedness and semantic incongruity become virtues qua surprise- and ambiguity-based cognitive challenge, as does prosodic regularity qua expectation-based perceptual ease
6. **Increased willlingness to invest additional interpretive effort** in the light of apparent semantic incongruity (*Exps. 1a/b*)

Our results clearly support Zwaan's model of literary comprehension. Particularly the neural and behavioral effects we observed in *Experiment 1b* corroborate the notion of a literary-comprehension control system (LCCS) that affects

the process and outcome of text comprehension. However, some extensions of Zwaan's basic model are necessary to account for the present findings. First, the contrastive evidence obtained in *Experiment 2* clearly indicates that there is not just one LCCS but distinct genre-appropriate modes of literary reading. Second, the EEG evidence obtained in *Experiment 1c* indicates that this genre-specific LCCS is active prior to reading proper and affects not only the process and outcome of text comprehension but already the anticipatory attentional state of the comprehender. Finally, the judgment data obtained in *Experiments 3a/b* indicates that the genre of poetry is associated with selected linguistic constructions; the behavioral evidence obtained in *Experiment 1b* and *Experiment 2* seems to indicate that the genre-specific LCCS incorporates schematic knowledge about prototypical and permissible units and relations at several linguistic levels of surface form and presumably at other levels of mental text representation as well. In this view, a (literary) genre corresponds to a set of constraints on potentially all levels of mental text representation, including several levels of surface form, the text base, the situation models and the text world that emerges from them (cf. Schauber & Spolsky, 1986), as well as the pragmatic model that specifies the communicative situation in terms of discourse participant roles. Whether and how a certain level of representation is constrained in any given genre depends on the units and relations that a reader has encountered repeatedly across texts and on the reliability of their occurrence across texts of the genre in question. Genre categorization then amounts to the activation of the memory schema that specifies these constraints and makes knowledge about prototypical and permissible units and relations available to the processing system. An active genre schema thus generates expectations already prior to reading, pre-activates expectable grammatical constructions (and presumably lexical cohorts as well), and leads the processing system to adjust information gathering strategies and eye-movement control. Literary genre schemata may co-determine phonetic text recoding and alter the interpretive strategies as well as the evaluative criteria applied to the phonological, grammatical and semantic representations constructed from written text.

Future directions

The present results were obtained in a population of experienced readers that are overwhelmingly literary non-experts. In terms of the cognitive theory of genre sketched in the Introduction, these readers have acquired basic literary genre schemata like “a poem” but likely with little differentiation into subgenres (Hanauer, 1996); nonetheless, we observed clear electrophysiological and behavioral effects of these basic text categories. It seems reasonable to assume that greater literary reading experience leads to richer schematic representations of literary genres (more represented knowledge and experience, more sub-schemata, more interrelations), and that generic differentiation also comes with formal training aimed at the development of such *literary competence* (Bierwisch, 1965; Culler, 1975; Ryan, 1979). Thus, the basic level of expertise is an intermediate step towards greater literary competence and presumably an intermediate developmental step in genre-specific reading modes. Therefore, we assume, following the logic of cognitive genre theory and based on various bits of available evidence (e.g., Funkhouser & O’Connell, 2013; Hanauer, 1995; Kowal et al., 1975; Peskin, 2010), that readers with greater literary expertise and competence will develop more differentiated and more pronounced genre-specific reading strategies. Specifically, we hypothesize that the observed behavioral differences between genre-specific reading strategies increase with the number of literary works known to a reader (i.e. the individual corpus) and the generalized literary categories extracted from them, e.g., genre or author schemata. A similar logic applies to our stimulus materials: strictly speaking, we cannot generalize most of our results to authentic texts that are rich in genre-specific features (*Exp. 3b* specifically addressed this issue). But if we can observe systematic differences even with constructed genre hybrids that are compatible with two different categorizations (e.g., verse or sentence, prose or poetry), then we should observe even more pronounced effects, i.e., a reinforcement of the genre-specific adjustments, if the constructed phonological, grammatical, and discourse-semantic representations resonate with the activated genre schema. Whether these hypotheses are correct needs to be assessed in future work.

The processing and evaluation of poetic forms, i.e., of poetry-specific formal and thematic schemata, constitutes an obvious area of investigation in literary psycholinguistics. The formal aspects of poetic rhyme, for instance, are very well described, but its cognitive effects on the process and outcome of text comprehenson remain relatively poorly understood. Here, we have focused on category-related effects and thus on poetry-specific reading at very early processing stages—none of our stimuli exceeded the length of a quatrain. Genre-specific reading dynamics that arise from the processing of larger poetic structures and of formal and thematic patterning across stanzas—while beyond the scope of the present thesis—clearly deserve further investigation.

A full understanding of genre-specific reading will have to be based on further systematic genre contrasts that allow for a more fine-grained mapping of literary comprehension, preferably in relation to readers' degree of literary competence. Differences between other literary (and non-literary) genres need to be addressed in future research so that, ultimately, single aspects of genre-specific reading behavior can be linked to their function for the respective reading strategy and to the formal, thematic and pragmatic text aspects that necessitate these strategic adjustments.

Finally, the available evidence for literary comprehension and evaluation strategies and for the processing of poetic forms stems from a handful of languages and literary traditions. If we wish to travel the path of psycholinguistic investigation to a better understanding of how literature works, we need to take the full diversity of verbal art into account, conduct cross-linguistic comparisons of literary processing phenomena and—similar to descriptive and theoretical metrics and poetics—broaden the range of literary traditions and poetic forms under investigation.

References

[1] Abraham, W. & Braunmüller, K. (1973). Towards a theory of style and metaphor. *Poetics*, 2(3), 103–148.

[2] Alter, A. L. (2013). The benefits of cognitive disfluency. *Current Directions in Psychological Science*, 22(6), 437–442.

[3] Altmann, U., Bohrn, I. C., Lubrich, O., Menninghaus, W., & Jacobs, A. M. (2014). Fact vs fiction—how paratextual information shapes our reading processes. *Social Cognitive and Affective Neuroscience*, 9(1), 22–29.

[4] Aristotle (1961). *Poetics. Transl. by S. H. Butcher*. New York, NY: Hill and Wang.

[5] Ashby, J., Yang, J., Evans, K. H., & Rayner, K. (2012). Eye movements and the perceptual span in silent and oral reading. *Attention, Perception, and Psychophysics*, 74(4), 634–640.

[6] Baayen, H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

[7] Baayen, H., Davidson, D. J., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412.

[8] Baldick, C. (2015). *The Oxford Dictionary of Literary Terms*. Oxford University Press.

[9] Barney, T. (1999). Readers as text processors and performers: A new formula for poetic intonation. *Discourse Processes*, 28(2), 155–167.

[10] Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge: Cambridge University Press.

[11] Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015a). *Parsimonious mixed models*. Available from arXiv:1506.04967 (stat.ME): https://arxiv.org/abs/1506.04967.

[12] Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015b). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.

[13] Beaugrande, R.-A. (1978). Information, expectation, and processing: On classifying poetic texts. *Poetics*, 7(1), 3–44.

[14] Benedek, M., Schickel, R. J., Jauk, E., Fink, A., & Neubauer, A. C. (2014). Alpha power increases in right parietal cortex reflects focused internal attention. *Neuropsychologia*, 56, 393–400.

[15] Benjamini, Y. & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1), 289–300.

[16] Bentin, S., Kutas, M., & Hillyard, S. A. (1993). Electrophysiological evidence for task effects on semantic priming in auditory word processing. *Psychophysiology*, 30(2), 161–9.

[17] Bierwisch, M. (1965). Poetik und Linguistik. In H. Kreuzer & R. Gunzenhäuser (Eds.), *Mathematik und Dichtung* (pp. 49–66). Munich: Nymphenburg.

[18] Blohm, S., Menninghaus, W., & Schlesewsky, M. (2017). Sentence-level effects of literary genre: Behavioral and electrophysiological evidence. *Frontiers in Psychology*, 8(1887).

[19] Boersma, P. & Weenink, D. (2017). *Praat: Doing phonetics by computer*. http://www.praat.org/.

[20] Bohn, K., Knaus, J., Wiese, R., & Domahs, U. (2013). The influence of rhythmic (ir)regularities on speech processing: Evidence from an ERP study on German phrases. *Neuropsychologia*, 51(4), 760–71.

[21] Bolinger, D. (1965). *Forms of English; accent, morpheme, order [papers, by] Dwight L. Bolinger*. Cambridge, MA: Harvard University Press.

[22] Bornkessel, I., Fiebach, C. J., & Friederici, A. D. (2004). On the cost of syntactic ambiguity in human language comprehension: An individual differences approach. *Cognitive Brain Research*, 21(1), 11–21.

[23] Bornkessel, I. & Schlesewsky, M. (2006). The extended argument dependency model: A neurocognitive approach to sentence comprehension across languages. *Psychological Review*, 113(4), 787–821.

[24] Bornkessel-Schlesewsky, I. & Schlesewsky, M. (2008). An alternative perspective on "semantic P600" effects in language comprehension. *Brain Research Reviews*, 59(1), 55–73.

[25] Bornkessel-Schlesewsky, I. & Schlesewsky, M. (2009). The role of prominence information in the real-time comprehension of transitive constructions: A cross-linguistic approach. *Language and Linguistics Compass*, 3(1), 19–58.

[26] Bostanov, V. & Kotchoubey, B. (2004). Recognition of affective prosody: Continuous wavelet measures of event-related brain potentials to emotional exclamations. *Psychophysiology*, 41(2), 259–268.

[27] Bower, G. H. & Bolton, L. S. (1969). Why are rhymes easy to learn? *Journal of Experimental Psychology*, 82(3), 453–461.

[28] Breen, M. (2018). Effects of metric hierarchy and rhyme predictability on word duration in The Cat in the Hat. *Cognition*, 174, 71–81.

[29] Breen, M. & Clifton, C., J. (2011). Stress matters: Effects of anticipated lexical stress on silent reading. *Journal of Memory and Language*, 64(2), 153–170.

[30] Bröggelwirth, J. (2007). *Ein rhythmisch-prosodisches Modell lyrischen Sprechstils [A rhythmic-prosodic model of poetic speech]*. Thesis, University of Bonn.

[31] Brown, C. & Hagoort, P. (1993). The processing nature of the N400: Evidence from masked priming. *Journal of Cognitive Neuroscience*, 5(1), 34–44.

[32] Brysbaert, M. & Vitu, F. (1998). Word skipping: Implications for theories of eye movement control in reading. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 125–147). Amsterdam: Elsevier.

[33] Bühler, K. (1934). *Sprachtheorie: Die Darstellungsfunktion der Sprache*. Jena: G. Fischer.

[34] Bullot, N. J. & Reber, R. (2013). The artful mind meets art history: Toward a psycho-historical framework for the science of art appreciation. *Behavioral and Brain Sciences*, 36(2), 123–137.

[35] Burkhardt, P. (2007). The P600 reflects cost of new information in discourse memory. *Neuroreport*, 18(17), 1851–4.

[36] Byers, P. P. (1979). A formula for poetic intonation. *Poetics*, 8(4), 367–380.

[37] Carpenter, P. A. & Just, M. A. (1983). What your eyes do while your mind is reading. In K. Rayner (Ed.), *Eye Movements in Reading: Perceptual and Language Processes* (pp. 275–307). New York: Academic Press.

[38] Carretié, L., Iglesias, J., & García, T. (1997a). A study on the emotional-processing of visual stimuli through event-related potentials. *Brain and Cognition*, 34(2), 207–217.

[39] Carretié, L., Iglesias, J., García, T., & Ballesteros, M. (1997b). N300, P300 and the emotional processing of visual stimuli. *Electroencephalography and Clinical Neurophysiology*, 103(2), 298–303.

[40] Castiglione, D. (2017). Difficult poetry processing: Reading times and the narrativity hypothesis. *Language and Literature*, 26(2), 99–121.

[41] Cermolacce, M., Scannella, S., Faugere, M., Vion-Dury, J., & Besson, M. (2014). "All that glitters is not ... alone". Congruity effects in highly and less predictable sentence contexts. *Clinical Neurophysiology*, 44(2), 189–201.

[42] Chen, Q., Zhang, J., Xu, X., Scheepers, C., Yang, Y., & Tanenhaus, M. K. (2016). Prosodic expectations in silent reading: ERP evidence from rhyme scheme and semantic congruence in classic chinese poems. *Cognition*, 154, 11–21.

[43] Chennu, S., Noreika, V., Gueorguiev, D., Blenkmann, A., Kochen, S., Ibáñez, A., Owen, A. M., & Bekinschtein, T. A. (2013). Expectation and attention in hierarchical auditory prediction. *The Journal of Neuroscience*, 33(27), 11194–11205.

[44] Chesnokova, A. & van Peer, W. (2016). Anyone came to live here some time ago: A cognitive semiotics approach to deviation as a foregrounding device. *Versus: Quaderni di studi semiotici*, 2016(1), 5–22.

[45] Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: The MIT Press.

[46] Chwilla, D. J., Brown, C. M., & Hagoort, P. (1995). The N400 as a function of the level of processing. *Psychophysiology*, 32(3), 274–85.

[47] Clifton, C., Fanselow, G., & Frazier, L. (2006). Amnestying superiority violations: Processing multiple questions. *Linguistic Inquiry*, 37(1), 51–68.

[48] Connolly, J. F. & Phillips, N. A. (1994). Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *Journal of Cognitive Neuroscience*, 6(3), 256–66.

[49] Coulson, S. & Van Petten, C. (2002). Conceptual integration and metaphor: An event-related potential study. *Memory and Cognition*, 30(6), 958–68.

[50] Cowart, W. (1997). *Experimental syntax: Applying objective methods to sentence judgments*. Thousand Oaks, CA: Sage.

[51] Craik, F. I. M. & Lockhart, R. S. (1972). Levels of processing – a framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11, 671–684.

[52] Culler, J. D. (1975). *Structuralist poetics: Structuralism, linguistics and the study of literature*. Ithaca, NY: Cornell University Press.

[53] Cureton, R. (1980). Poetic syntax and aesthetic form. *Style*, 14(4), 318–340.

[54] Davies, M. (2008). *The Corpus of Contemporary American English (COCA): 520 million words, 1990-present*. http://corpus.byu.edu/coca/.

[55] Del Percio, C., Le Pera, D., Arendt-Nielsen, L., Babiloni, C., Brancucci, A., Chen, A., De Armas, L., Miliucci, R., Restuccia, D., Valeriani, M., & Rossini, P. (2006). Distraction affects frontal alpha rhythms related to expectancy of pain: An EEG study. *Neuroimage*, 31(3), 1268–1277.

[56] DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word preactivation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, 8, 1117–21.

[57] Dryer, M. S. (2013). Order of genitive and noun. In M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

[58] Egner, T., Monti, J. M., & Summerfield, C. (2010). Expectation and surprise determine neural population responses in the ventral visual stream. *The Journal of Neuroscience*, 30(49), 16601–16608.

[59] Einstein, G. O., McDaniel, M. A., Owen, P. D., & Coté, N. C. (1990). Encoding and recall of texts: The importance of material appropriate processing. *Journal of Memory and Language*, 29(5), 566–581.

[60] Enkvist, N. E. (1973). *Linguistic stylistics*, volume 5 of *Janua linguarum: Series critica*. The Hague: Mouton.

[61] Fabb, N. (1997). *Linguistics and literature: Language in the verbal arts of the world*, volume 12 of *Blackwell textbooks in linguistics*. Oxford: Blackwell.

[62] Fabb, N. (2001). Weak monosyllables in iambic verse and the communication of metrical form. *Lingua*, 111(11), 771–790.

[63] Fabb, N. (2010). Is literary language a development of ordinary language? *Lingua*, 120(5), 1219–1232.

[64] Fabb, N. (2015). *What is poetry? Language and memory in the poems of the world*. Cambridge: Cambridge University Press.

[65] Fanselow, G. (2009). Unrealised possibilities. *Snippets*, 20, 18–22.

[66] Fanselow, G. & Frisch, S. (2006). Effects of processing difficulty on judgments of acceptability. In G. Fanselow, C. Féry, M. Schlesewsky, & R. Vogel (Eds.), *Gradience in grammar: Generative perspectives* (pp. 291–316). Oxford: Oxford University Press.

[67] Fant, G., Kruckenberg, A., & Nord, L. (1991). Stress patterns and rhythm in the reading of proseand poetry with analogies to music performance. In J. Sundberg, L. Nord, & R. Carlson (Eds.), *Music, Language, Speech and Brain: Proceedings of an International Symposium at the Wenner-Gren Center, Stockholm, 5–8 September 1990* (pp. 380–407). London: Macmillan Education UK.

[68] Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, 44(4), 491–505.

[69] Federmeier, K. D. & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language*, 41(4), 469–495.

[70] Fischer, M. H., Carminati, M. N., Stabler, J., & Roberts, A. (2003). Eye movements during poetry and prose reading. In *13th European Conference on Eye Movements, Dundee*.

[71] Fish, S. E. (1970). Literature in the reader: Affective stylistics. *New Literary History*, 2(1), 123–162.

[72] Fish, S. E. (1973). How ordinary is ordinary language? *New Literary History*, 5(1), 41–54.

[73] Fishelov, D. (1998). The institutional definition of poetry: Some heretical thoughts. *Empirical Studies of the Arts*, 16(1), 5–13.

[74] Fitzgerald, C. M. (2007). An optimality treatment of syntactic inversions in English verse. *Language Sciences*, 29(2–3), 203–217.

[75] Fodor, J. D. (1998). Learning to parse? *Journal of Psycholinguistic Research*, 27, 285–319.

[76] Fox, J. & Weisberg, S. (2011). *An R companion to applied regression*. Thousand Oaks, CA: Sage.

[77] Foxe, J. J. & Snyder, A. C. (2011). The role of alpha-band brain oscillations as a sensory suppression mechanism during selective attention. *Frontiers in Perception Science*, 2.

[78] Frenzel, S., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2015). Two routes to actorhood: Lexicalized potency to act and identification of the actor role. *Frontiers in Psychology*, 6(1).

[79] Friederici, A. D. & Frisch, S. (2000). Verb-argument structure processing: The role of verb-specific and argument-specific information. *Journal of Memory and Language*, 43, 476–507.

[80] Frisch, S. & Schlesewsky, M. (2001). The N400 reflects problems of thematic hierarchizing. *Neuroreport*, 12(15), 3391–4.

[81] Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1456), 815–815.

[82] Friston, K. (2012). Prediction, perception and agency. *International Journal of Psychophysiology*, 83(2), 248–52.

[83] Friston, K. & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1521), 1211.

[84] Funkhouser, L. & O'Connell, D. C. (2013). Temporal aspects of poetry readings by authors and adults. *Bulletin of the Psychonomic Society*, 12(5), 390–392.

[85] Ghosh, V. E. & Gilboa, A. (2014). What is a memory schema? A historical perspective on current neuroscience literature. *Neuropsychologia*, 53(1), 104–114.

[86] Gibbs, R. W., Kushner, J. M., & Mills, W. R. (1991). Authorial intentions and metaphor comprehension. *Journal of Psycholinguistic Research*, 20(1), 11–30.

[87] Graesser, A. C. (1981). *Prose comprehension beyond the word*. New York, NY: Springer.

[88] Graf, L. K. M. & Landwehr, J. R. (2015). A dual-process perspective on fluency-based aesthetics: The pleasure-interest model of aesthetic liking. *Personality and Social Psychology Review*, 19(4), 395–410.

[89] Gregory, R. L. (1980). Perceptions as hypotheses. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 290(1038), 181–197.

[90] Hagoort, P., Baggio, G., & Willems, R. M. (2009). Semantic unification. In M. S. Gazzaniga (Ed.), *The Cognitive Neurosciences* (pp. 819–836). Cambridge, MA: MIT Press.

[91] Haider, H. (1988). Die Struktur der deutschen Nominalphrase [The structure of the German noun phrase]. *Zeitschrift für Sprachwissenschaft*, 7(1), 32–59.

[92] Haider, H. (2010). *The syntax of German*. Cambridge syntax guides. Cambridge: Cambridge University Press.

[93] Hakemulder, J. F. (2004). Foregrounding and its effect on readers' perception. *Discourse Processes*, 38(2), 193–218.

[94] Halliday, M. & Hasan, R. (1976). *Cohesion in English*. London: Longman.

[95] Hanauer, D. (1995). Literary and poetic text categorization judgements. *Journal of Literary Semantics*, 24(3), 187–209.

[96] Hanauer, D. (1996). Integration of phonetic and graphic features in poetic text categorization judgements. *Poetics*, 23(5), 363–380.

[97] Hanauer, D. (1997). Poetic text processing. *Journal of Literary Semantics*, 26(3), 157–172.

[98] Hanauer, D. (1998a). The genre-specific hypothesis of reading: Reading poetry and encyclopedic items. *Poetics*, 26(2), 63–80.

[99] Hanauer, D. (1998b). Reading poetry: An empirical investigation of formalist, stylistic, and conventionalist claims. *Poetics Today*, 19(4), 565–580.

[100] Händel, B. F., Haarmeier, T., & Jensen, O. (2011). Alpha oscillations correlate with the successful inhibition of unattended stimuli. *Journal of Cognitive Neuroscience*, 23(9), 2494–2502.

[101] Hasko, S., Bruder, J., Bartling, J., & Schulte-Korne, G. (2012). N300 indexes deficient integration of orthographic and phonological representations in children with dyslexia. *Neuropsychologia*, 50(5), 640–54.

[102] Haupt, F. S., Schlesewsky, M., Roehm, D., Friederici, A. D., & Bornkessel-Schlesewsky, I. (2008). The status of subject–object reanalyses in the language comprehension architecture. *Journal of Memory and Language*, 59(1), 54–96.

[103] Havránek, B. (1964). The functional differentiation of the standard language. In P. L. Garvin (Ed.), *A Prague School reader on esthetics, literary structure, and style* (pp. 3–16). Washington, DC: Georgetown University Press.

[104] Hemforth, B. (1993). *Kognitives Parsing: Repräsentation und Verarbeitung sprachlichen Wissens*. St. Augustin: Infix.

[105] Henrich, K., Alter, K., Wiese, R., & Domahs, U. (2014). The relevance of rhythmical alternation in language processing: An ERP study on English compounds. *Brain and Language*, 136, 19–30.

[106] Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, 93(4), 411–428.

[107] Hoffstaedter, P. (1987). Poetic text processing and its empirical investigation. *Poetics*, 16(1), 75–91.

[108] Hofmeister, J., Heller, D., & Radach, R. (1999). The return sweep in reading. In W. Becker, H. Deubel, & T. Mergner (Eds.), *Current Oculomotor Research: Physiological and Psychological Aspects* (pp. 349–357). Boston, MA: Springer.

[109] Hofmeister, P., Jaeger, T. F., Arnon, I., Sag, I., & Snider, N. (2013). The source ambiguity problem: Distinguishing the effects of grammar and processing on acceptability judgments. *Language and Cognitive Processes*, 28, 48–87.

[110] Hoorn, J. (1996). Psychophysiology and literary processing: Erps to semantic and phonological deviations in reading small verses. In R. Kreuz & M. MacNealy (Eds.), *Empirical Approaches to Literature and Aesthetics* (pp. 339–358). Norwood, NJ: Ablex.

[111] Hopkins, G. M. (1959). Poetic diction. In H. House & G. Storey (Eds.), *The journals and papers of Gerard Manley Hopkins* (pp. 84–85). London: Oxford University Press.

[112] Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, 50(3), 346–63.

[113] Ibn Fāris, A. (10th ct. AD). Reproving mistakes in poetry. In A. Sanni (Ed.), *A Fourth Century Contribution to Literary Theory: Ibn Fāris's Treatise on Poetic Licenses* (pp. 14–18). (A. Sanni, trans.).

[114] Idsardi, W. J. & Raimy, E. (2005). *Remarks on language play (Ms.)*. University of Delaware; Swarthmore College.

[115] Ingarden, R. (1931). *Das literarische Kunstwerk: Eine Untersuchg aus d. Grenzgebiet d. Ontologie, Logik u. Literaturwissenschaft [The literary work of art: An investigation of the borderlines of ontology, logic, and literary theory]*. Tübingen: Niemeyer.

[116] Jacobs, A. M. (2015a). Neurocognitive poetics: Methods and models for investigating the neuronal and cognitive-affective bases of literature reception. *Frontiers in Human Neuroscience*, 9(186).

[117] Jacobs, A. M. (2015b). Towards a neurocognitive poetics model of literary reading. In R. M. Willems (Ed.), *Cognitive Neuroscience of Natural Language Use* (pp. 135–159). Cambridge: Cambridge University Press.

[118] Jakobson, R. (1960). Closing statement: Linguistics and poetics. In T. A. Sebeok (Ed.), *Style in Language* (pp. 350–377). MIT Press.

[119] Jensen, O. & Mazaheri, A. (2010). Shaping functional architecture by oscillatory alpha activity: Gating by inhibition. *Frontiers in Human Neuroscience*, 4, 186.

[120] Jespersen, O. (1933). *Linguistica: Selected papers in English, French and German*. Copenhagen: Levin.

[121] Kant, I. (1781). *Critique of pure reason*. (N. Kemp Smith, trans.) London: Macmillan and Co., 1963, originally published, 1781, 1st edition.

[122] Kelly, S. P., Lalor, E. C., Reilly, R. B., & Foxe, J. J. (2006). Increases in alpha oscillatory power reflect an active retinotopic mechanism for distracter suppression during sustained visuospatial attention. *Journal of Neurophysiology*, 95(6), 3844–3851.

[123] Kentner, G. (2012). Linguistic rhythm guides parsing decisions in written sentence comprehension. *Cognition*, 123(1), 1–20.

[124] Kentner, G. (2015). Problems of prosodic parallelism: A reply to wiese and speyer (2015). *Linguistics*, 53(5), 1233.

[125] Kentner, G. & Vasishth, S. (2016). Prosodic focus marking in silent reading: Effects of discourse context and rhythm. *Frontiers in Psychology*, 7(319).

[126] Kien, J. & Kemp, A. (1994). Is speech temporally segmented? Comparison with temporal segmentation in behavior. *Brain and Language*, 46(4), 662–682.

[127] Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95(2), 163–182.

[128] Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge, MA: Cambridge University Press.

[129] Kiparsky, P. (1975). Stress, syntax, and meter. *Language*, 51(3), 576–616.

[130] Kisler, T., Reichel, U., & Schiel, F. (2017). Multilingual processing of speech via web services. *Computer Speech and Language*, 45, 326–347.

[131] Klimesch, W. (2012). Alpha-band oscillations, attention, and controlled access to stored information. *Trends in Cognitive Sciences*, 16(12), 606–617.

[132] Klimesch, W., Fellinger, R., & Freunberger, R. (2011). Alpha oscillations and early stages of visual encoding. *Frontiers in Psychology*, 2(118).

[133] Knoop, C. A., Wagner, V., Jacobsen, T., & Menninghaus, W. (2016). Mapping the aesthetic space of literature “from below”. *Poetics*, 56, 35–49.

[134] Kok, P., Jehee, J. F. M., & de Lange, F. P. (2012). Less is more: Expectation sharpens representations in the primary visual cortex. *Neuron*, 75(2), 265–270.

[135] Koops van ’t Jagt, R., Hoeks, J. C., Dorleijn, G. J., & Hendriks, P. (2014). Look before you leap: How enjambment affects the processing of poetry. *Scientific Study of Literature*, 4(1), 3–24.

[136] Kounios, J. & Holcomb, P. J. (1994). Concreteness effects in semantic processing: ERP evidence supporting dual-coding theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(4), 804–23.

[137] Kowal, S., O’Connell, D. C., O’Brien, E. A., & Bryant, E. T. (1975). Temporal aspects of reading aloud and speaking: Three experiments. *The American Journal of Psychology*, 88(4), 549–569.

[138] Kraxenberger, M. & Menninghaus, W. (2016). Emotional effects of poetic phonology, word positioning and dominant stress peaks in poetry reading. *Scientific Study of Literature*, 6(2), 298–313.

[139] Kriukova, O. & Mani, N. (2016). Processing metrical information in silent reading: An ERP study. *Frontiers in Psychology*, 7(1432).

[140] Kruckenberg, A. & Fant, G. (1993). Iambic versus trochaic patterns in poetry reading. In B. Granström & L. Nord (Eds.), *Proceedings of Nordic Prosody VI, Stockholm* (pp. 123–136). Almqvist and Wiksell International.

[141] Kuehnast, M., Wagner, V., Wassiliwizky, E., Jacobsen, T., & Menninghaus, W. (2014). Being moved: Linguistic representation and conceptual structure. *Frontiers in Psychology*, 5.

[142] Kutas, M. & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (erp). *Annual Review of Psychology*, 62, 621–47.

[143] Kutas, M. & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207(4427), 203–5.

[144] Kutas, M. & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, 307(5947), 161–3.

[145] Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmertest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 26.

[146] Lau, E. F., Namyst, A., Fogel, A., & Delgado, T. (2016). A direct comparison of N400 effects of predictability and incongruity in adjective-noun combination. *Collabra*, 2, 1–19.

[147] Lea, R. B., Rapp, D. N., Elfenbein, A., Mitchel, A. D., & Romine, R. S. (2008). Sweet silent thought: Alliteration and resonance in poetry comprehension. *Psychological Science*, 19(7), 709–16.

[148] Leech, G. N. (1969). *A linguistic guide to English poetry*. Harlow: Longmans.

[149] Lenerz, J. (1979). Dichterische Freiheit, linguistisch betrachtet. In K. Grubmüller (Ed.), *Befund und Deutung: Zum Verhältnis von Empirie und Interpretation in Sprach-und Literaturwissenschaft* (pp. 80–95). Tübingen: Niemeyer.

[150] Levin, S. R. (1962). *Linguistic Structures in Poetry*. The Hague: Mouton.

[151] Levin, S. R. (1965). Internal and external deviation in poetry. *Word*, 21(2), 225–237.

[152] Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, 49(4), 764–766.

[153] Luo, Y. & Zhou, X. (2010). ERP evidence for the online processing of rhythmic pattern during chinese sentence reading. *Neuroimage*, 49(3), 2836–49.

[154] MacWhinney, B., Bates, E., & Kliegl, R. (1984). Cue validity and sentence interpretation in english, german, and italian. *Journal of Verbal Learning and Verbal Behavior*, 23(2), 127–150.

[155] Magne, C., Gordon, R. L., & Midha, S. (2010). Influence of metrical expectancy on reading words: An ERP study. In *Proceedings of the Fifth International Conference on Speech Prosody* Chicago, IL.

[156] Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco, New York: W.H. Freeman.

[157] Martindale, C. (1973). Approximation to natural language, grammaticalness, and poeticality. *Poetics*, 3(1), 21–25.

[158] Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing type i error and power in linear mixed models. *Journal of Memory and Language*, 94, 305–315.

[159] May, E. S., Butz, M., Kahlbrock, N., Hoogenboom, N., Brenner, M., & Schnitzler, A. (2012). Pre- and post-stimulus alpha activity shows differential modulation with spatial attention during the processing of pain. *NeuroImage*, 62(3), 1965–1974.

[160] McDaniel, M. A., Einstein, G. O., Dunay, P. K., & Cobb, R. E. (1986). Encoding difficulty and memory: Toward a unifying theory. *Journal of Memory and Language*, 25(6), 645–656.

[161] McGlone, M. S. & Tofighbakhsh, J. (1999). The Keats heuristic: Rhyme as reason in aphorism interpretation. *Poetics*, 26(4), 235–244.

[162] McGlone, M. S. & Tofighbakhsh, J. (2000). Birds of a feather flock conjointly (?): Rhyme as reason in aphorisms. *Psychological Science*, 11(5), 424–8.

[163] McGowan, V. A. & Reichle, E. D. (2018). The “risky“ reading strategy revisited: New simulations using e-z reader. *Quarterly Journal of Experimental Psychology*, 71(1), 179–189.

[164] Menninghaus, W., Bohrn, I. C., Altmann, U., Lubrich, O., & Jacobs, A. M. (2014). Sounds funny? Humor effects of phonological and prosodic figures of speech. *Psychology of Aesthetics, Creativity, and the Arts*, 8(1), 71–76.

[165] Menninghaus, W., Bohrn, I. C., Knoop, C. A., Kotz, S. A., Schlotz, W., & Jacobs, A. M. (2015). Rhetorical features facilitate prosodic processing while handicapping ease of semantic comprehension. *Cognition*, 143, 48–60.

[166] Menninghaus, W., Wagner, V., Wassiliwizky, E., Jacobsen, T., & Knoop, C. A. (2017). The emotional and aesthetic powers of parallelistic diction. *Poetics*, 63, 47–59.

[167] Meutsch, D. (1986). Mental models in literary discourse: Towards the integration of linguistic and psychological levels of description. *Poetics*, 15(3), 307–331.

[168] Meutsch, D. (1987). *Literatur verstehen: Eine empirische Studie*. Braunschweig: Vieweg.

[169] Meutsch, D. & Schmidt, S. J. (1985). On the role of conventions in understanding literary texts. *Poetics*, 14(6), 551–574.

[170] Miall, D. S. & Kuiken, D. (1994). Foregrounding, defamiliarization, and affect: Response to literary stories. *Poetics*, 22(5), 389–407.

[171] Minsky, M. (1975). A framework for representing knowledge. In P. H. Winston (Ed.), *The Psychology of Computer Vision* (pp. 211–277). New York: McGraw-Hill.

[172] Mukařovský, J. (1964). Standard language and poetic language. In P. L. Garvin (Ed.), *A Prague School reader on esthetics, literary structure, and style* (pp. 17–30). Washington, DC: Georgetown University Press.

[173] Müller, S. (2003). Mehrfache Vorfeldbesetzung. *Deutsche Sprache*, 31(1), 29–62.

[174] Nieuwland, M. S. & Van Berkum, J. J. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience*, 18(7), 1098–111.

[175] Nord, L., Kruckenberg, A., & Fant, G. (1990). Some timing studies of prose, poetry and music. *Speech Communication*, 9(5–6), 477–483.

[176] Obermeier, C., Kotz, S. A., Jessen, S., Raettig, T., von Koppenfels, M., & Menninghaus, W. (2016). Aesthetic appreciation of poetry correlates with ease of processing in event-related potentials. *Cognitive, Affective and Behavioral Neuroscience*, 16(2), 362–373.

[177] Obermeier, C., Menninghaus, W., von Koppenfels, M., Raettig, T., Schmidt-Kassow, M., Otterbein, S., & Kotz, S. A. (2013). Aesthetic and emotional effects of meter and rhyme in poetry. *Frontiers in Psychology*, 4.

[178] Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, 9(1), 97–113.

[179] Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. (2011). Fieldtrip: Open source software for advanced analysis of meg, eeg, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, 2011, doi:10.1155/2011/156869.

[180] Osterhout, L. & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, 31, 785–806.

[181] O'Sullivan, N., Davis, P., Billington, J., Gonzalez-Diaz, V., & Corcoran, R. (2015). "Shall I compare thee": The neural basis of literary awareness, and its benefits to cognition. *Cortex*, 73, 144–157.

[182] Paltridge, B. (1995). Working with genre: A pragmatic perspective. *Journal of Pragmatics*, 24, 393–406.

[183] Pechmann, T., Uszkoreit, H., Engelkamp, J., & Zerbst, D. (1994). *Word order in the German middlefield*. Report 43, University of the Saarland.

[184] Peskin, J. (1998). Constructing meaning when reading poetry: An expert-novice study. *Cognition and Instruction*, 16(3), 235–263.

[185] Peskin, J. (2007). The genre of poetry: High school students' expectations and interpretive operations. *English in Education*, 41(3), 20–36.

[186] Peskin, J. (2010). The development of poetic literacy during the school years. *Discourse Processes*, 47(2), 77–103.

[187] Pilkington, A. (2000). *Poetic effects: A relevance therory perspective*. Pragmatics and beyond. Amsterdam: Benjamins.

[188] Praamstra, P., Meyer, A. S., & Levelt, W. J. (1994). Neurophysiological manifestations of phonological processing: Latency variation of a negative ERP component timelocked to phonological mismatch. *Journal of Cognitive Neuroscience*, 6(3), 204–19.

[189] Praamstra, P. & Stegeman, D. F. (1993). Phonological effects on the auditory N400 event-related brain potential. *Cognitive Brain Research*, 1(2), 73–86.

[190] R-Core-Team (2016). *R: A language and environment for statistical computing*. https://www.R-project.org/.

[191] Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372–422.

[192] Rayner, K. & Duffy, S. A. (1988). On-line comprehension processes and eye movements in reading. In M. Daneman, G. E. MacKinnon, & T. G. Waller (Eds.), *Reading research: Advances in theory and practice* (pp. 13–66). New York, NY: Academic Press.

[193] Rayner, K. & McConkie, G. W. (1976). What guides a reader's eye movements? *Vision Research*, 16(8), 829–837.

[194] Rayner, K., Reichle, E. D., Stroud, M. J., Williams, C. C., & Pollatsek, A. (2006). The effect of word frequency, word predictability, and font difficulty on the eye movements of young and older readers. *Psychology of Aging*, 21(3), 448–65.

[195] Rice, C. (1997). Ranking components: The grammar of poetry. In G. Booij & J. van de Weijer (Eds.), *Phonology in Progress – Progress in Phonology* (pp. 321–332). The Hague: Holland Academic Graphics.

[196] Riffaterre, M. (1960). Stylistic context. *Word*, 16(2), 207–218.

[197] Roehm, D., Bornkessel-Schlesewsky, I., Rösler, F., & Schlesewsky, M. (2007). To predict or not to predict: Influences of task and strategy on the processing of semantic relations. *Journal of Cognitive Neuroscience*, 19(8), 1259–1274.

[198] Rosenblatt, L. M. (1978). *The reader, the text, the poem: The transactional theory of the literary work*. Carbondale, Ill.: Southern Illinois University Press.

[199] Rosenblatt, L. M. (1988). *Writing and Reading: The Transactional Theory*. Report 13, National Center for the Study of Writing.

[200] Rösler, F., Pechmann, T., Streb, J., Röder, B., & Hennighausen, E. (1998). Parsing of sentences in a language with varying word order: Word-by-word variations of processing demands are revealed by event-related brain potentials. *Journal of Memory and Language*, 38(2), 150–176.

[201] Rothermich, K., Schmidt-Kassow, M., & Kotz, S. A. (2012). Rhythm's gonna get you: Regular meter facilitates semantic sentence processing. *Neuropsychologia*, 50(2), 232–44.

[202] Rubin, D. C. (1995). *Memory in oral traditions: The cognitive psychology of epic, ballads, and counting-out rhymes*. New York: Oxford University Press.

[203] Rubin, D. C., Ciobanu, V., & Langston, W. (1997). Children's memory for counting-out rhymes: A cross-language comparison. *Psychonomic Bulletin and Review*, 4(3), 421–424.

[204] Rugg, M. D. (1984a). Event-related potentials and the phonological processing of words and non-words. *Neuropsychologia*, 22(4), 435–43.

[205] Rugg, M. D. (1984b). Event-related potentials in phonological matching tasks. *Brain and Language*, 23(2), 225–40.

[206] Rugg, M. D. & Barrett, S. E. (1987). Event-related potentials and the interaction between orthographic and phonological information in a rhyme-judgment task. *Brain and Language*, 32(2), 336–61.

[207] Rumelhart, D. E. (1975). Notes on a schema for stories. In D. G. Bobrow & A. Collins (Eds.), *Representation and Understanding. Studies in Cognitive Science* (pp. 211–236). New York: Academic Press.

[208] Rumelhart, D. E. (1977). Understanding and summarizing brief stories. In D. LaBerge & J. Samuels (Eds.), *Basic Processes in Reading: Perception and Comprehension* (pp. 1–39). Hillsdale, NJ: Lawrence Erlbaum.

[209] Rumelhart, D. E. (1980). Schemata: The building blocks of cognition. In R. J. Spiro (Ed.), *Theoretical Issues in Reading Comprehension* book section 2, (pp. 33–58). Hillsdale, NJ: Lawrence Erlbaum.

[210] Rumelhart, D. E. & Ortony, A. (1977). The representation of knowledge in memory. In R. C. Anderson, R. J. Spiro, & W. E. Montague (Eds.), *Schooling and the acquisition of knowledge* (pp. 99–135). Hillsdale, NJ: Erlbaum.

[211] Ryan, M.-L. (1979). Toward a competence theory of genre. *Poetics*, 8(3), 307–337.

[212] Sanni, A. (1993). A fourth century contribution to literary theory: Ibn Fāris's treatise on poetic licenses. *Journal of Arabic Literature*, 24(1), 11–20.

[213] Schank, R. C. & Abelson, R. P. (1975). Scripts, plans, and knowledge. In *Proceedings of the Proceedings of the 4th International joint conference on Artificial intelligence - Volume 1*, volume 1 (pp. 151–157).

[214] Schank, R. C. & Abelson, R. P. (1977). *Scripts, Plans, Goals and Understanding: An Inquiry into Human Knowledge Structures*. Hillsdale, NJ: Lawrence Erlbaum.

[215] Schauber, E. & Spolsky, E. (1986). *The bounds of interpretation: Linguistic theory and literary text*. Stanford, CA: Stanford University Press.

[216] Schlüter, J. (2005). *Rhythmic grammar: The influence of rhythm on grammatical variation and change in English*, volume 46 of *Topics in English linguistics*. Berlin: Mouton de Gruyter.

[217] Schröger, E., Marzecová, A., & SanMiguel, I. (2015). Attention and prediction in human audition: A lesson from cognitive psychophysiology. *European Journal of Neuroscience*, 41(5), 641–664.

[218] Schumacher, P. B. & Avrutin, S. (2011). Register affects language comprehension: ERP evidence from article omission in newspaper headlines. *Journal of Neurolinguistics*, 24(3), 304–319.

[219] Schumacher, P. B. & Hung, Y.-C. (2012). Positional influences on information packaging: Insights from topological fields in german. *Journal of Memory and Language*, 67(2), 295–310.

[220] Schütze, C. T. (1996). *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Chicago, IL: University of Chicago Press.

[221] Selkirk, E. (1984). *Phonology and Syntax: The Relation between Sound and Structure*. Cambridge, MA: The MIT Press.

[222] Selkirk, E. (1996). The prosodic structure of function words. In J. L. Morgan & K. Demuth (Eds.), *Signal to Syntax: Boot-strapping from Speech to Grammar in Early Acquisition* (pp. 187–214). Mahwah, NJ: Lawrence Erlbaum Associates.

[223] Shelley, P. B. (1840). A defence of poetry. In M. Shelley (Ed.), *Essays, Letters from abroad, Translations and Fragments*, volume 1. London: Moxon.

[224] Šhklovsky, V. (1965). Art as technique. In L. T. Lemon & M. J. Reis (Eds.), *Russian formalist criticism: Four essays* (pp. 3–24). Lincoln, NE: University of Nebraska Press.

[225] Smith, B. H. (1968). *Poetic closure: A study of how poems end*. Chicago, IL: University of Chicago Press.

[226] Sohoglu, E., Peelle, J. E., Carlyon, R. P., & Davis, M. H. (2012). Predictive top-down integration of prior knowledge during speech perception. *The Journal of Neuroscience*, 32(25), 8443–8453.

[227] Sparks, J. R. & Rapp, D. N. (2010). Discourse processing—examining our everyday language experiences. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1(3), 371–381.

[228] Speyer, A. (2004). Competing constraints on Vorfeldbesetzung in German. In B. Shaer, W. Frey, & C. Maienborn (Eds.), *Proceedings of the Dislocated Elements Workshop: ZAS Berlin, November 2003*, volume 35 of *ZAS papers in linguistics* (pp. 519–541). Berlin: Zentrum für Allgemeine Sprachwissenschaft, Sprachtypologie und Universalienforschung.

[229] Spiro, R. J. (1977). Remembering information from text: The state of schema approach. In R. C. Anderson, R. J. Spiro, & M. W. E. (Eds.), *Schooling and the Acquisition of Knowledge* (pp. 137–166). Hillsdale, NJ: Lawrence Erlbaum.

[230] Spironelli, C. & Angrilli, A. (2007). Influence of phonological, semantic and orthographic tasks on the early linguistic components N150 and N350. *International Journal of Psychophysiology*, 64(2), 190–198.

[231] Sprouse, J. (2013). Acceptability judgments. *Oxford Bibliographies in Linguistics*.

[232] Steen, G. (1999). Genres of discourse and the definition of literature. *Discourse Processes*, 28(2), 109–120.

[233] Steinhauer, K. & Friederici, A. D. (2001). Prosodic boundaries, comma rules, and brain responses: The closure positive shift in erps as a universal marker for prosodic phrasing in listeners and readers. *Journal of Psycholinguistic Research*, 30(3), 267–95.

[234] Stolterfoht, B., Friederici, A. D., Alter, K., & Steube, A. (2007). Processing focus structure and implicit prosody during reading: Differential ERP effects. *Cognition*, 104(3), 565–90.

[235] Summerfield, C. & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences*, 13(9), 403–409.

[236] Sussman, E., Winkler, I., Huotilainen, M., Ritter, W., & Näätänen, R. (2002). Top-down effects can modify the initially stimulus-driven auditory organization. *Cognitive Brain Research*, 13(3), 393–405.

[237] Szczepaniak, R. (2010). Während des Flug(e)s/des Ausflug(e)s: German short and long genitive endings between norm and variation. In A. Lenz & A. Plewnia (Eds.), *Grammar between norm and variation* (pp. 103–126). Frankfurt: Peter Lang.

[238] Szewczyk, J. M. & Schriefers, H. (2011). Is animacy special? ERP correlates of semantic violations and animacy violations in sentence processing. *Brain Research*, 1368, 208–21.

[239] Thorndyke, P. W. & Yekovich, F. R. (1980). A critique of schema-based theories of human story memory. *Poetics*, 9(1), 23–49.

[240] Thorne, J. (1988). What is a poem? In W. van Peer (Ed.), *The Taming of the Text: Explorations in Language, Literature and Culture,* (pp. 280–291). London: Routledge.

[241] Thorpe, S., D'Zmura, M., & Srinivasan, R. (2012). Lateralization of frequency-specific networks for covert spatial attention to auditory stimuli. *Brain Topography*, 25(1), 39–54.

[242] Thut, G., Nietzel, A., Brandt, S. A., & Pascual-Leone, A. (2006). α-band electroencephalographic activity over occipital cortex indexes visuospatial attention bias and predicts visual target detection. *The Journal of Neuroscience*, 26(37), 9494.

[243] Tillmann, B. & Dowling, W. J. (2007). Memory decreases for prose, but not for poetry. *Memory and Cognition*, 35(4), 628–639.

[244] Todorovic, A., van Ede, F., Maris, E., & de Lange, F. P. (2011). Prior expectation mediates neural adaptation to repeated sounds in the auditory cortex: An meg study. *The Journal of Neuroscience*, 31(25), 9118–9123.

[245] Tremblay, A. Ransijn, J. (2015). *LMERConvenienceFunctions: Model selection and post-hoc analysis for (G)LMER models*, volume R package version 2.10. https://CRAN.R-project.org/package=LMERConvenienceFunctions.

[246] Uszkoreit, H. (1986). *Constraints on order*. Stanford, CA: Center for the Study of Language and Information.

[247] van Albada, S. J. & Robinson, P. A. (2013). Relationships between electroencephalographic spectral peaks across frequency bands. *Frontiers in Human Neuroscience*, 7, 56.

[248] van Dijk, T. A. & Kintsch, W. (1983). *Strategies of Discourse Comprehension*. New York: Academic Press.

[249] van Peer, W. (1986). *Stylistics and psychology: Investigations of foregrounding*. Croom Helm linguistics series. London: Croom Helm.

[250] van Peer, W. (1990). The measurement of metre: Its cognitive and affective functions. *Poetics*, 19(3), 259–275.

[251] van Peer, W. (2007). Introduction to foregrounding: A state of the art. *Language and Literature*, 16(2), 99–104.

[252] Vančová, H. (2014). Eye movements during the reading of literary text. *Language, Individual and Society*, 8, 150–157.

[253] Vaughan-Evans, A., Trefor, R., Jones, L., Lynch, P., Jones, M. W., & Thierry, G. (2016). Implicit detection of poetic harmony by the naive brain. *Frontiers in Psychology*, 7(1859).

[254] Verdaasdonk, H. (1982). Conceptions of literature as frames? *Poetics*, 11(1), 87–104.

[255] Viehoff, R. (1995). Literary genres as cognitive schemata. In G. Rusch (Ed.), *Empirical Approaches to Literature* (pp. 72–76). Siegen: LUMIS.

[256] Vogel, R., van de Vijver, R., Kotz, S., Kutscher, A., & Wagner, P. (2015). Function words in rhythmic optimisation. In R. Vogel & R. van de Vijver (Eds.), *Rhythm in Cognition and Grammar: A Germanic Perspective* (pp. 255–276). Berlin, Boston: De Gruyter Mouton.

[257] Wagner, P. (2012). Meter specific timing and prominence in German poetry and prose. In O. Niebuhr (Ed.), *Understanding Prosody – The Role of Context, Function and Communication* (pp. 219–236). New York, NY: De Gruyter.

[258] 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.

[259] West, W. C. & Holcomb, P. J. (2000). Imaginal, semantic, and surface-level processing of concrete and abstract words: An electrophysiological investigation. *Journal of Cognitive Neuroscience*, 12(6), 1024–37.

[260] Widmann, A., Schroger, E., & Maess, B. (2015). Digital filter design for electrophysiological data: A practical approach. *Journal of Neuroscientific Methods*, 250, 34–46.

[261] Wiese, R. (2000). *The Phonology of German*. The phonology of the world's languages. Oxford: Oxford University Press.

[262] Wiese, R. (2016). Prosodic parallelism-comparing spoken and written language. *Frontiers in Psychology*, 7(1598).

[263] Wiese, R. & Speyer, A. (2015). Prosodic parallelism explaining morphophonoogical alternation in German. *Linguistics*, 53, 525–559.

[264] Wimsatt, W. K. J. (1944). One relation of rhyme to reason: Alexander Pope. *Modern Language Quarterly*, 5, 323–338.

[265] Wolff, S., Schlesewsky, M., Hirotani, M., & Bornkessel-Schlesewsky, I. (2008). The neural mechanisms of word order processing revisited: Electrophysiological evidence from Japanese. *Brain and Language*, 107(2), 133–57.

[266] Wordsworth, W. (1802). *Lyrical ballads*. Abingdon: Routledge.

[267] Youmans, G. (1983). Generative tests for generative meter. *Language*, 59(1), 67–92.

[268] Zwaan, R. A. (1991). Some parameters of literary and news comprehension: Effects of discourse-type perspective on reading rate and surface structure representation. *Poetics*, 20(2), 139–156.

[269] Zwaan, R. A. (1993). *Aspects of literary comprehension: A cognitive approach*, volume 29 of *Utrecht publications in general and comparative literature*. Amsterdam: Benjamins.

[270] Zwaan, R. A. (1994). Effect of genre expectations on text comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(4), 920–933.

[271] Zwaan, R. A. (1996). Toward a model of literary comprehension. In B. K. Britton & A. C. Graesser (Eds.), *Models of understanding text* (pp. 241–256). Mahwah, NJ: Erlbaum.

Zusammenfassung

Die vorliegende Dissertation untersucht Prozesse und Mechanismen der Sprachwahrnehmung, die charakteristisch für die Rezeption literarischer Texte sind. Sie berichtet eine Reihe von Studien, die mittels behavioraler und elektrophysiologischer Messmethoden den Einfluss von Gattungskonzeptionen auf die Verarbeitung und Evaluation sprachlicher Reize untersuchen. Was diese Studien verbindet, ist die systematische Gegenüberstellung gedichtspezifischer und alltagssprachlicher bzw. prosaspezifischer Verarbeitungs- und Evaluationsroutinen unter Beibehaltung des sprachlichen Materials. Was diese Studien unterscheidet, sind die untersuchten Aspekte des Rezeptionsprozesses und die dazu verwendeten, der Psycholinguistik entlehnten Methoden.

Aus linguistischer Perspektive sind diese Studien der experimentellen Pragmatik zuzuordnen, denn sie leisten einen Beitrag zu unserem Verständnis kontextabhängiger Sprachverstehensprozesse, indem sie deutlicher machen als zuvor, dass und wie der Äußerungskontext, hier die Textsorte bzw. -gattung, unterschiedliche Aspekte der Sprachverarbeitung mitbestimmt. Aus literaturwissenschaftlicher Sicht fördern sie somit unser Verständnis kognitiver Konzeptionen literarischer Gattungen.

Kapitel 1 berichtet zwei Studien, die mithilfe systematisch gesammelter Leserintuitionen (*Experiment 1a*) und ereigniskorrelierter Hirnpotentiale (EKPs; *Experiment 1b*) untersuchen, welchen Einfluss Gedicht- und Verskonzeptionen auf das Echtzeitverstehen und die Beurteilung einzelner Sätze haben, die gattungstypische formale und semantische Merkmale aufweisen.

Kapitel 2 berichtet weitere Analysen der in *Experiment 1b* gesammelten EEG-Daten. Mittels Zeit-Frequenz- und EKP-Analysen untersucht *Experiment 1c* den Einfluss von Gattungszuschreibung auf antizipatorische Aufmerksamkeit vor dem Lesen und auf die frühe Echtzeitverarbeitung geschriebener Sprache.

Kapitel 3 berichtet eine Studie, die mithilfe kombinierter Blickbewegungsmessungen und Sprachaufnahmen (*Experiment 2*) untersucht, wie Gattungszuschrei-

bung (Gedicht vs. literarische Prosa) das (Vor)lesen unbekannter Texte beeinflusst, und so behaviorale und akustische Marker dieser literarischen Lesemodi identifiziert. Im Gegensatz zu früheren Untersuchungen zielt diese Studie explizit darauf ab, distinktive Dynamiken gattungspezifischen Lesens zu erfassen.

Kapitel 4 berichtet zwei Studien, die sich systematisch gesammelter Leserintuitionen bedienen, um den Einfluss syntaktischer und prosodischer Variablen auf die grammatische und literarisch-ästhetisch Evaluation einzelner Verse (*Experiment 3a*) und Sätze (*Experiment 3b*) zu untersuchen. Diese Studien zeigen, dass erfahrene Leser die Textsorte Gedicht mit spezifischen grammatischen Strukturen und Merkmalen verbinden. Ich argumentiere, dass das Auftreten dieser Textmerkmale charakteristische Verarbeitungsschritte erfordert bzw. erleichtert, die von Lesern als spezifisch poetische Qualitäten des Textes und der Leseeerfahrung wahrgenommen werden.

Kapitel 5 fasst die Resultate der berichteten Studien zusammen, bezieht sie auf einzelne Stufen des Rezeptionsprozesses und diskutiert die Grenzen ihrer Generalisierbarkeit. Das Kapitel schließt mit einem Ausblick auf mögliche zukünftige Forschungsfelder literarischer Psycholinguistik.

Auf der Grundlage der gesammelten Beobachtungen ergibt sich folgende Liste gedichtspezifischer Anpassungen der Sprachverarbeitung und -evaluation.
Die Konzeption der Textsorte Gedicht...

1. ...beeinflusst die Erwartungen und die Aufmerksamkeit der Leser bereits vor dem Lesen
2. ...bewirkt jedoch *nicht*, dass Leser prosodischen Rekurrenzen gesteigerte Aufmerksamkeit schenken
3. ...führt zu gattungsspezifischen Anpassungen des Leseverhaltens und der Blickbewegungsroutinen
4. ...bestimmt Strategien zur gattungsadäquaten phonetischen Realisierung sprachlicher Reize

5. ...beeinflusst zwar *nicht* die frühe Verarbeitung (scheinbarer) semantischer Inkongruenz, veranlasst Leser jedoch, sowohl während des Lesens als auch danach größeren Interpretationsaufwand zu betreiben, um zu einer kohärenten Bedeutungsrepräsentation zu gelangen

6. ...beeinflusst *nicht* die strategische Nutzung verarbeiteter prosodischer Regelmäßigkeiten (Metrum) beim Lesen

7. ...macht Leser resilienter gegenüber gattungstypischen historischen Wortformen, die ohne Kenntnis der Gattung kurzzeitig die verfügbaren kognitiven Ressourcen während des Lesens reduzieren

8. ...beinhaltet gattungsspezifische Evaluationskriterien für grammatische und semantische Merkmale sprachlicher Reize

Gemeinsam betrachtet zeigen diese Resultate deutlich, dass literarische Gattungen selbst in der Konzeption literarischer Laien mit ausgewählten sprachlichen Konstruktionen und Merkmalen verknüpft sind, sowie mit adäquaten Anpassungen der Sprachverarbeitung und -evaluation. Diese Anpassungen betreffen sämtliche Stufen des literarischen Rezeptionsprozesses: die Aufmerksamkeit des Lesers *vor* dem Lesen, die Verarbeitungsroutinen *während* des Lesens, und die Evaluationskriterien für sprachliche Reize *nach* dem Lesen.

Die in dieser Arbeit untersuchte Leserpopulation ist eine studentische: junge Muttersprachler in Hochschulausbildung, die durch den Schulunterricht zumindest grundlegend mit der behandelten Gattung vertraut, aber großteils literarische Laien mit diversen akademischen Hintergründen und literarischen Interessen sind. Basierend auf der kognitiven Gattungstheorie von Steen (1999) und gestützt durch vereinzelte Befunde aus verwandten Untersuchungen (z.B., Funkhouser & O'Connell, 2013; Hanauer, 1995; Kowal et al., 1975; Peskin, 2010) argumentiere ich, dass die beobachteten Anpassungen des Lese- und Evaluationsverhaltens einer Zwischenstufe gattungsspezifischer Sprachverarbeitung entsprechen, und dass Leser mit wachsender Leseerfahrung differenziertere Gattungskonzeptionen und zunehmend deutlich voneinander zu unterscheidende literarische Lesemodi entwickeln.