

Comprehending irony in text: evidence from scanpaths

Henri Olkonemi, Diane Mézière & Johanna K. Kaainen

To cite this article: Henri Olkonemi, Diane Mézière & Johanna K. Kaainen (16 Oct 2023): Comprehending irony in text: evidence from scanpaths, *Discourse Processes*, DOI: [10.1080/0163853X.2023.2266964](https://doi.org/10.1080/0163853X.2023.2266964)

To link to this article: <https://doi.org/10.1080/0163853X.2023.2266964>



© 2023 The Author(s). Published with license by Taylor & Francis Group, LLC.



Published online: 16 Oct 2023.



Submit your article to this journal [↗](#)



Article views: 231



View related articles [↗](#)



View Crossmark data [↗](#)

Comprehending irony in text: evidence from scanpaths

Henri Olkonieni ^{a,b}, Diane Mézière ^b, and Johanna K. Kaakinen ^{b,c}

^aDivision of Psychology, Faculty of Education and Psychology, University of Oulu; ^bDepartment of Psychology and Speech-language Pathology, University of Turku; ^cINVEST Research Flagship, University of Turku



ABSTRACT

Eyetracking studies have shown that readers reread ironic phrases when resolving their meaning. Moreover, it has been shown that the timecourse of processing ironic meaning is affected by reader's working memory capacity (WMC). Irony is a context-dependent phenomenon but using traditional eye-movement measures it is difficult to analyze processing beyond sentence-level. A promising method to study individual differences in irony processing at the paragraph-level is scanpath analysis. In the present experiment, we analyzed whether individual differences in WMC are reflected in scanpaths during reading ironic stories by combining data from two previous eye-tracking studies ($N = 120$). The results revealed three different reading patterns: fast-and-linear reading, selective reading, and nonselective rereading. The readers predominantly used the fast-and-linear reading pattern for ironic and literal stories. However, readers were less likely to use the nonselective rereading pattern with ironic than literal texts. The reading patterns for ironic stories were modulated by WMC. Results showed that scanpaths captured differences missed by standard measures, showing it to be a valuable tool to study individual differences in irony processing.

Introduction

In John Green's novel *The Fault in Our Stars* (2012), one of the characters, Isaac, has been diagnosed with eye cancer. Isaac is having a conversation with his surgeon, who says "Well, the good news is that you won't be deaf." Isaac responds: "Thank you for explaining that my eye cancer isn't going to make me deaf. I feel so fortunate that an intellectual giant like yourself would deign to operate on me." (Green, 2012, p. 15). While reading the dialogue, it becomes clear that Isaac is not happy about the surgeon's comment and that Isaac does not literally mean what he said but quite the opposite. Isaac's response is an example of verbal irony, which is a form of figurative language in which something opposite is intended than what is literally said (Attardo, 2000). Irony is typically used to criticize *something* or *someone* (the latter is also called sarcasm), which is what Isaac meant to do with his response (e.g., Kreuz & Link, 2002). Irony is a context-dependent phenomenon (Spotorno & Noveck, 2019). In the prior example, the situation and the surgeon's statement create a context which Isaac's comment contrasts. Understanding this contrast is necessary for interpreting Isaac's comment as ironic. Irony is also rarely a property of a single word, but the property of the whole phrase (e.g., Olkonieni & Kaakinen, 2021; Spotorno & Noveck, 2019); this is also the case in the previous example.

Eye-tracking is a methodology that allows detailed analysis of the time-course of processing different parts of text while not posing extra demands on the reader (e.g., Rayner, 1998), and in recent years it has been used to study irony processing (e.g., Filik et al., 2018; Filik & Moxey, 2010; Kaakinen et al., 2014; Olkonieni et al., 2016, 2023; Turcan & Filik, 2016). Despite irony being a phrase-

CONTACT Henri Olkonieni  henri.olkonieni@oulu.fi  Faculty of Education and Psychology, University of Oulu, P.O.Box 2000, Oulu 90014, Finland

© 2023 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

level phenomenon that does not exist without context, previous eye movement studies on processing written irony have focused on studying the reading time of the target phrase or specific words within the phrase (see Olkonieni & Kaakinen, 2021 for review). However, focusing on the target phrase does not provide information on how irony affects processing at the whole text-level. This is important from the comprehension point-of-view, as readers need to integrate the ironic phrase with the context to resolve its meaning, which might involve rereading all or certain parts of the text. One reason behind this focus in previous research has been a lack of suitable analytical tools to examine reading patterns at the text-level. Relatively recently, Von der Malsburg and Vasishth (2011) proposed an algorithm for scanpath analysis, which can be used to identify reading patterns from eye movement data by analyzing how readers transition between different parts of text. Scanpath analysis is also suitable for examining individual differences in the eye movement patterns (e.g., Von der Malsburg & Vasishth, 2013).

In the current study, we combine data from two of our previous eye-tracking studies (Kaakinen et al., 2014 Exp 2; Olkonieni et al., 2016) and, for the first time, use scanpath methodology to study the processing and comprehension of irony. This allows us to identify individual differences in the reading patterns used to resolve irony at the whole text-level. Specifically, we are interested in the effects of readers' working memory capacity (WMC), which has been shown to affect irony processing at the phrase-level (see, e.g., Kaakinen et al., 2014; Olkonieni et al., 2016).

Processing of written irony

The theories on irony comprehension make different assumptions about the need of activating the literal meaning of a phrase before the ironic meaning could be accessed. According to the classical *standard pragmatic view* (e.g., Grice, 1975) the discrepancy between the literal meaning of a phrase and the context in which it appears is central to irony processing. First, a reader makes a literal interpretation of the phrase. Second, a discrepancy between the literal interpretation and the context is detected. Last, the reader seeks an alternative interpretation and comprehends that the phrase is ironic. Because the comprehension is seen as a serial three-step-process, comprehension of irony is expected to be more difficult and take more time than comprehending a literal phrase.

Later theories on irony comprehension make the same assumption when specific conditions are satisfied. They assume that an ironic phrase should be harder to comprehend and take more time to process when the phrase is not familiar as irony (*the graded salience view*, Giora, 2003) and when the preceding context does not provide cues about forthcoming irony (*the direct access view*, Gibbs, 1994). In the present study, we combined data from two of our previous studies (Kaakinen et al., 2014; Olkonieni et al., 2016) that examined processing of nonfamiliar ironic phrases in story contexts that did not provide support or cues for an ironic interpretation.

According to the different theoretical views, these types of ironic phrases should be harder to comprehend and slower to process than literal phrases. Previous eye-tracking studies support this assumption (see Olkonieni & Kaakinen, 2021), as they have shown that in this kind of setting (i.e., irony is not familiar or strongly supported by context) processing of ironic phrases takes longer than processing of their literal counterparts (e.g., Filik & Moxey, 2010). This slowdown in processing has been typically seen as increased looking back to ironic target phrases after they have been already read once (e.g., Kaakinen et al., 2014; Olkonieni et al., 2016). In addition, readers are more likely to return to the preceding context from ironic than from literal phrases (e.g., Olkonieni & Kaakinen, 2021).

Understanding what kind of reading patterns are associated with successfully resolving irony is important for developing the theories of irony comprehension. The previous eye movement studies have focused on fixation time measures on ironic (vs. literal) target phrases, and sometimes on one or two contextual sentences (see Olkonieni & Kaakinen, 2021). However, these types of analyses provide little information about the reading patterns across the text. For example, returning to the context and to the ironic target phrase may reflect selective rereading of the interpretation-relevant parts of the text, or it could reflect more random, nonselective rereading of the whole text (e.g., Hyönä &

Nurminen, 2006; Hyönä et al., 2002). Selective rereading of relevant parts of text might indicate a successful strategy for resolving the ironic meaning of a phrase, whereas nonselective rereading might be related to comprehension difficulties (e.g., Hyönä et al., 2002; León et al., 2019). Moreover, linear reading, in which the reader does not return to the preceding context at all, may indicate that the reader does not even notice irony, or that they resolve the ironic meaning efficiently during first-pass reading. In the present study, we used scanpath analysis to examine what kind of reading patterns emerge during reading of ironic and literal texts, and whether certain eye movement patterns are associated with better comprehension of irony than others.

Moreover, it is important to know if the different eye-movement patterns reflect successful comprehension of irony or a comprehension failure (see Ferreira & Yang, 2019). Only some previous eye-tracking studies report measures of irony comprehension (e.g., Au-Yeung et al., 2015; Kaakinen et al., 2014; Olkonieni et al., 2016, 2023). Thus, in general, we cannot be certain to what extent the previous eye-tracking findings reflect irony comprehension. Of the studies reporting comprehension accuracy, only a few have reported a relationship between comprehension and reading times (Olkonieni et al., 2023; Olkonieni, Johander, et al., 2019). For example, Olkonieni et al. (2023) showed that better irony comprehension accuracy was related to increased first-pass reading times of the target phrase and the region immediately following the phrase. This is in line with more general findings showing an increase in reading times when text is more difficult to understand, and that this increase is related to better comprehension (see e.g., Rayner, 2009 for review). However, increased reading effort does not automatically indicate better comprehension. A study by Olkonieni, Johander et al. (2019) showed that a higher probability to look-back from the ironic target phrase to the preceding context was associated with poorer irony comprehension.

Although these two studies provide some information on how the processing of texts containing irony is related to irony comprehension, results from these studies should be interpreted with caution. First, Olkonieni et al. (2023) studied children's irony comprehension and the stories used in the experiment were suitable for children. Adults served as a control group in the study and their results might not be generalizable for texts that are more relevant for adult readers. Second, Olkonieni, Johander, et al. (2019) used a masking paradigm to study the role of look-backs in the processing of written irony. The masking used in the study disrupted normal reading, which might affect the generalizability of the results.

Role of WMC in irony processing

There are individual differences in processing and comprehending ironic phrases, and previous studies suggest that WMC is associated with the efficiency of resolving irony in text (Kaakinen et al., 2014; Olkonieni et al., 2016; Olkonieni, Johander, et al., 2019; see also Antoniou & Milaki, 2021; Godbee & Porter, 2013). It has been suggested that readers with a low WMC have problems with inhibiting the automatically activated salient, typically literal, meaning (Giora, 1999). Consequently, difficulties with inhibiting the literal meaning result in deficits with integrating the ironic interpretation of the phrase with the context. To overcome these deficits, readers with lower WMC need to use compensatory strategies, for example, slow down their reading or do more rereading of the text (Walczyk & Taylor, 1996).

Some of the newer theories on irony comprehension also take explicitly into account individual differences in irony comprehension (*the parallel-constraint-satisfaction framework*, Pexman, 2008; *the predictive coding theory on irony*; Fabry, 2021). For example, the *predictive coding theory* on irony (Fabry, 2021) is based on the broader assumption that all biological organisms aim to minimize prediction error (Friston et al., 2006), and individual differences in perception, cognition, and emotional abilities affect the efficiency of this minimization.

From the perspective of irony comprehensions, this means that as the use of irony is not often expected, the priors are set to predict a literal interpretation. When the reader encounters an ironic phrase, a prediction error occurs. This requires then a corrective process to form a more suitable

interpretation, and the priors might be adjusted to minimize prediction error in the future. From an individual differences point of view, for example, individuals' higher WMC aids them to make better predictions during reading by allowing them to maintain relevant priors in accessible states for the subsequent interpretation process (Trapp et al., 2021), but also to recover from prediction errors. For example, higher WMC has been associated with more strategic rereading behavior (Hyönä et al., 2002).

Previous studies on the processing of written irony have shown that high WMC is related to increased first-pass reading times on ironic phrases (Kaakinen et al., 2014; Olkonieni et al., 2016), and faster reading times after the target phrase (i.e., spillover region; Kaakinen et al., 2014). In addition, individuals with higher WMC are faster in making ironic interpretations, after observing videoed dialogs containing an ironic comment (Antoniou & Milaki, 2021). In contrast, low WMC readers are more likely to rely on later rereading (i.e., look-backs) of the ironic phrase (Olkonieni et al., 2016; Olkonieni, Johander, et al., 2019). Similar results have been obtained for metaphors. In their eye-tracking study, Columbus et al. (2015, Exp. 1) explored the role of context and executive control (WMC and executive control has been shown to be tightly related, see e.g., McCabe et al., 2010) in processing written metaphors.

The results showed that readers with higher executive control had longer first-pass reading times, whereas readers with low executive control were more likely to do later rereading (i.e., regressing back to context word). By and large, these results are in line with theories that take into account individual differences (Fabry, 2021; Pexman, 2008; see also Giora, 1999), as it seems that readers with higher WMC would be better at suppressing the salient literal meaning and/or maintaining interpretation relevant priors in accessible state, as they can start processing the intended ironic meaning already during the earliest stages of processing. However, Olkonieni, Strömberg, et al. (2019) showed that when the ironic phrase appears in a relatively short context (1 or 2 sentences), the WMC effect is not observed. This suggests that WMC impacts irony processing only when readers must inhibit the irrelevant context that is not needed for the interpretation and there is heavier load on the working memory functions. In the present study, we examined whether WMC is associated with specific eye movement patterns during reading of ironic and literal texts, which included relevant and irrelevant contextual information.

Scanpath analyses of eye movement data

The problem in many of the eye-tracking studies on written irony (see Olkonieni & Kaakinen, 2021 for a review) is that they typically report fixation time measures calculated on regions of interest (typically single words, phrases, or sentences; see Rayner, 1998), which do not capture global eye-movement behavior and reading patterns beyond these regions of interests. Although such measures are informative, they are not optimal for describing the time-course of reading of the whole phrase (e.g., Kaakinen, 2017), which is required for interpreting ironic meaning. For example, imagine reading the phrase “What wonderful weather for a walk” in a text describing a scene with pouring rain. Here, the word “wonderful” defines the valence of the phrase, but the reader does not know whether “wonderful” is used in its literal sense until they have reached the end of the phrase. Moreover, fixation time measures provide only restricted information about the transitions between the different parts of text or global reading strategies used while reading longer passages of text. One approach that has been used to study transition patterns and global reading strategies is *scanpath analysis*.

A scanpath is a sequence of fixations made within a given time-series, as defined by their location (x- and y-coordinates) and duration. As such, scanpaths capture the overall pattern of gaze behavior during reading and can be used to investigate global reading patterns of sentences or whole passages of text. Critically, scanpaths have been shown to capture effects that are not necessarily caught by word-level measures (Mézière et al., 2022; Parshina et al., 2022; Von der Malsburg & Vasishth, 2011, 2013). In an early study, Von der Malsburg and Vasishth (2013) used scanpaths to identify the processing

patterns of readers when they encountered garden path sentences and considered the effect of WMC on which pattern readers used. They identified three main reading patterns, which varied according to where readers regressed to and from when processing garden-path sentences: The first pattern showed re-reading of the whole sentence, the second pattern showed re-readings of the critical area prior to the ambiguity, and the third pattern showed re-readings of the ambiguity from the spillover region.

Critically, these distinct patterns could not have been identified with standardly used word-level measures, showing that scanpaths can capture features of eye-movement behavior not caught by word-level measures. In addition, they found that readers' re-reading behavior and the extent to which they used those three patterns interacted with their WMC, showing that scanpaths are also useful to investigate the role of individual differences in processing patterns used during reading (see also Mézière et al., 2022; Parshina et al., 2022). Hence, scanpaths provide an ideal tool to investigate readers' eye movement behavior during processing of written irony, as well as the role of individual differences such as WMC or comprehension accuracy in reading literal and ironic passages of text.

The present study

In the present study we combined data from two of our previous eye-tracking studies (Kaakinen et al., 2014, Exp 2; Olkonieni et al., 2016) to investigate scanpaths during reading of texts containing literal and ironic comments. To this end, we calculated scanpaths starting from the first fixation made on the target phrase, to identify eye-movement patterns associated with resolving the meaning (literal or ironic) of the phrase. In addition, we investigated how WMC and comprehension of the target phrase are associated with the different scanpaths. More specifically, we had four research questions: First, we explored what kind of scanpaths occur during reading of literal and ironic texts, and whether these scanpaths differ between these text types. Based on the previous studies, we expect that immediate returns to the context would indicate sensitivity to irony and attempts to resolve it. Second, we examined how the different scanpaths differ with respect to the *key* eye movement measures (i.e., first-pass reading time, look-frogs, and look-backs) observed in the previous eye-tracking studies on written irony (see Olkonieni & Kaakinen, 2021 for review). Third, we examined individual differences related to WMC in the scanpaths for literal and ironic texts. We expect that the increase in WMC would be related to the likelihood of using a more linear reading pattern. Last, we investigated whether scanpath patterns reflecting immediate returns to context are related to better comprehension of irony.

Methods

Participants

We combined data from two of our previous experiments (Kaakinen et al., 2014, Exp 2; Olkonieni et al., 2016). Both experiments had 60 participants (49 women in both experiments), making a total of 120 participants (98 women). The participants were between the ages of 18 and 45, were native Finnish speakers (the language studied in both experiments), the majority of them University of Turku students, and all of them had normal or corrected-to-normal vision. In both experiments, participants gave written informed consent before the experiment.

Apparatus

In Kaakinen et al. (2014, Exp 2) eye movements were recorded with an EyeLink II or a desktop-mounted EyeLink 1000 eye-tracker system (SR Research Ltd.), using 500 Hz sampling frequency. In Olkonieni et al. (2016) EyeLink 1000 was used, using 1000 Hz sampling frequency. The eye-movement registration was always done monocularly, typically for the right eye. In both experiments, the stimuli were presented on a 21" CRT screen with a screen resolution of 1,024 × 768 pixels and

85 Hz refresh rate. Participants were seated 70 cm from the screen, and with EyeLink 1000 a chin rest was used to stabilize the head.

Materials

In both experiments materials were written in Finnish, the native language of the participants, and the experimental stories were similarly structured. In the Kaakinen et al. study (Kaakinen et al., 2014, Exp 2) the materials consisted of 24 experimental stories (and 12 filler stories); in the Olkonieni et al. (2016) study there were 30 experimental stories. In both experiments, font size used was Courier New, font size 14, and line-height 2.5. Each story had an ironic and a literal version (in Olkonieni et al., 2016, there were also metaphors but they were excluded from the present analyses), and each participant only saw either the literal or ironic version. The presentation of the story versions (literal vs. ironic) were counterbalanced across participants, and the presentation order of the stories was randomized. In both studies, each story was followed by two questions: a text memory question and an inference question designed to measure how a participant had understood the meaning of the target phrase. Based on participants' responses to the questions, mean text memory and inference question accuracy was calculated for each participant. An example of an English translation of an experimental story and the text memory and inference questions are presented in Table 1. The original stories (in Finnish) are available at <https://osf.io/5xs2z/>. Accuracies for answering the text memory and inference questions for literal and ironic texts in the two studies are presented in Table 2.

Table 1. Examples of experimental texts (translated from Finnish).

Text region	Version	Study	
		Kaakinen et al. (2014)	Olkonieni et al. (2016)
Introduction		Old high-school friends had planned to have a weekend picnic together. Anne had asked one of her friends to stay at her place for the night before the picnic. On Saturday morning Anne woke up and looked out from the window.	Kevin and Peter are having lunch at a school café.
Critical Context	Ironic	It was raining heavily and the street was flooding.	Peter piles up his plate, but at the table he's just pushing his food from one side of his plate to another.
	Literal	The sun was shining and the weather was warm.	Peter piles up his plate and his belly crumbles loudly.
Target Sentence		"What great weather for a picnic!"	"Well, you're hungry today!"
Spillover Region		Anne said and her friend came to the window too.	Kevin says to him.
Ending		They decided to spend the day in the house watching TV.	He suggests to Peter, that they could go to the city afterwards.
Memory Question		Was a friend staying at Anne's?	Were Kevin and Peter at the grill?
Inference Question		Was Anne happy about the weather?	Did Kevin think that Peter was hungry?

Note. English translations of the stimuli are available upon request from the first author.

Table 2. Proportion of correct answers to text memory and inference questions.

Question	Study	Literal		Irony	
		M	SD	M	SD
Text memory	Kaakinen et al. (2014)	.89	.08	.89	.11
	Olkonieni et al. (2016)	.90	.09	.93	.09
Inference	Kaakinen et al. (2014)	.93	.07	.83	.12
	Olkonieni et al. (2016)	.96	.08	.90	.11

The text materials were pretested for various factors. In Kaakinen et al. (2014, Exp 2) study, 50 participants who did not take part in the actual experiment rated how well the target phrases fitted to the context (1 = not at all, 5 = very well), and how ironic the target phrases were in the story context (1 = not at all ironic, 5 = very ironic). Both ironic ($M = 3.66$, $SD = 0.52$) and literal ($M = 4.34$, $SD = 0.34$) versions of the stories were rated as fitting to the context. Moreover, the target phrases were rated to be more ironic when presented in the ironic context ($M = 4.63$, $SD = 0.27$) than in the nonironic context ($M = 1.61$, $SD = 0.44$), indicating that the ironic meaning was not inherently salient for the target phrases.

In Olkonemi et al. (2016), the stories used were pretested for their familiarity (i.e., how often they had previously seen/heard the utterances in literal/ironic use; 0 = never, 10 = very often; $N = 25$ raters), and naturalness of the target phrase (0 = unnatural, 10 = very natural; $N = 34$ raters) by participants who did not take part in the eye-tracking experiment. Ironic and literal phrases were rated to be equally familiar ($M_{\text{Literal}} = 3.03$, $SD_{\text{Literal}} = 1.51$, $M_{\text{Ironic}} = 2.89$, $SD_{\text{Ironic}} = 2.31$), and natural ($M_{\text{Literal}} = 6.89$, $SD_{\text{Literal}} = 1.69$, $M_{\text{Ironic}} = 7.52$, $SD_{\text{Ironic}} = 1.22$).

Reading span test

A Finnish version of the *reading span* test (Daneman & Carpenter, 1980; Kaakinen & Hyönä, 2007) was used to measure verbal WMC in both studies. In the test, participants read aloud sets of unrelated sentences presented on a computer screen. After each set, they were asked to recall the sentence-final words of each sentence they read within that set. The test started with a set of two sentences, and the number of sentences increased if the participant was able to recall the final words of at least one of the trials in the set. The test was preceded by a practice session including three sets of two sentences. The test was scored for the total number of correctly recalled final words (scores vary between 0 and 81 points). Average score in Kaakinen et al. (2014) was 30.78 ($SD = 14.38$), and in Olkonemi et al. (2016), 27.13 ($SD = 13.04$).

Data analysis

The analysis was carried out in three main steps. First, we calculated scanpaths and ran a cluster analysis for each dataset separately to identify the global reading patterns that participants used when reading texts containing ironic and literal statements. A scanpath is a sequence of fixations generated from the location of the fixations (x- and y-coordinates on the screen) and their duration (in ms) and captures the global pattern of gaze behavior during reading, including reading (and fixation) times, skipping, and re-reading. Examples of scanpaths during text reading are provided in Figure 1, illustrating differences in both whole-text reading measures (e.g., trial reading time) and reading strategies (e.g., frequency and target of regressions). Second, the datasets were merged, and we used linear regression to investigate differences in reading behavior on three areas of interest (context, target sentence, spillover region). Third, we used multinomial regression models to investigate the relationship between scanpaths and irony processing, and considered the roles of WMC and comprehension. The analysis was conducted using R statistical software (Version 4.1.2; R Core Team, 2021). The data and the analysis code are shared via Open Science Framework (<https://osf.io/5xs2z/>). The original studies or the analysis plan of the present study were not preregistered.

The first step of the analysis was to generate and cluster scanpaths to identify the global reading patterns that participants used when reading passages of text that contained ironic and literal statements. This part of the analysis was carried out in the same way for the two datasets (i.e., Kaakinen et al., 2014, Exp 2; Olkonemi et al., 2016) separately, and follows the same steps used in previous studies investigating reading patterns and individual differences using scanpaths (e.g., Mézière et al., 2022; Parshina et al., 2022; Von der Malsburg & Vasisht, 2013).

The scanpaths were calculated starting from the target sentence which was either ironic or literal. For each item, dissimilarity between each pair of scanpaths was calculated using the *scasim* function

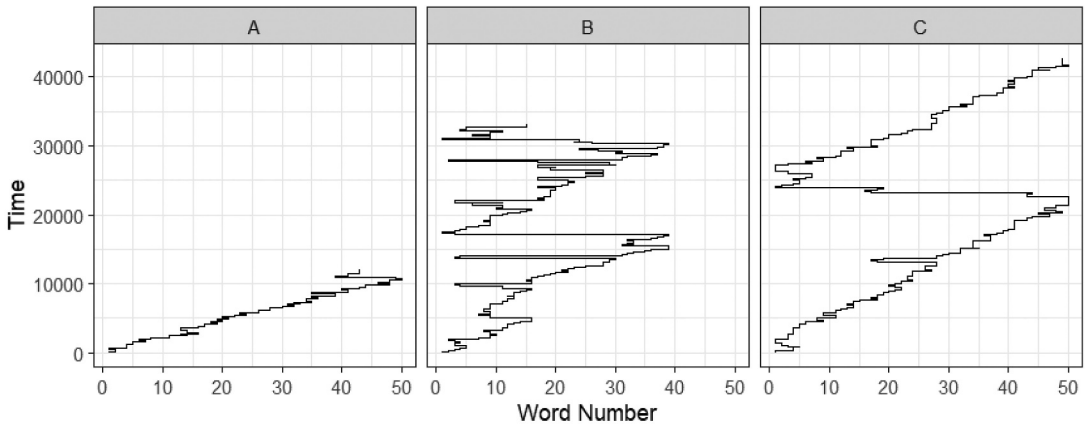


Figure 1. Example of scanpaths during text reading. *Note.* Figure 1 provides example of scanpaths calculated from text reading data and illustrates that scanpaths capture differences in global reading patterns such as total trial reading times (e.g., example A shows shorter trial reading time), as well as target and frequency of re-reading behavior (e.g., little to no re-reading in A vs. extensive re-readings in B and re-reading of the whole text once in C).

from the *scanpath* package (Von der Malsburg, 2018). Put simply, this measure can be thought of as the amount of time readers spent looking at different parts of the text. The dissimilarity scores were then used to calculate maps of scanpath space using the *isoMDS* function from the *MASS* package (Venables & Ripley, 2002). On these maps, similarity is represented by distance such that scanpaths that are similar to each other are close on the map and scanpaths that are dissimilar are far away. The number of dimensions to fit the maps was set at 3 to maximize the amount of variance explained across items while avoiding overfitting (see Von der Malsburg & Vasishth, 2011). The maps were then used to cluster scanpaths using Gaussian mixture modeling with the *mclust* package (Scrucca et al., 2016). The aim of the clustering was to identify distinct patterns of gaze behavior. The number of clusters was first allowed to vary freely across items. Based on the results of the clustering across items on both datasets and visualization of the clusters, the number of clusters was set at three for consistency across items and datasets. As the similar three scanpath clusters were identified in the two datasets, the datasets were then combined for the following analyses.

Next, we ran linear mixed effect models with the *lme4* package (Bates et al., 2015) to investigate differences in reading times between the three clusters identified in the scanpath analyses. We ran models comparing reading times on three areas of interest: the critical context, the target sentence, and the spillover region. Specifically, we ran models with the following 7 dependent variables: 1) first-pass reading of target, 2) first-pass re-reading of target, 3) look-backs to target, 4) look-backs from target, 5) first-pass reading of spillover region, 6) re-readings on the spillover region, and 7) look-backs to the context. For each reading time measure, we ran a model with cluster as predictor and random intercepts for participants and items. All reading time measures were skewed and, consequently, log-transformed prior to running the models.

Finally, we ran multinomial regression models with the *mclgit* package (Elff, 2022) to investigate whether clusters were related to text type, WMC, and comprehension. We ran four models with cluster membership as the dependent variable, with Pattern A as the baseline for comparison. We ran models with the following predictors: 1) text type, 2) text type \times WMC, 3) text type \times memory question accuracy, and 4) text type \times inference question accuracy. We also included random intercepts and slopes for text types for participants in all models. As the number of observations per cluster for each item was highly limited, we could not include random slopes of text type per item. Therefore, we only included random intercepts for item in model 1, and in models 2–4 we included random intercepts, and slopes for WMC, memory accuracy, and inference accuracy respectively. All the continuous predictors were scaled, and text type was dummy coded (literal = 0, ironic = 1). As we were also

interested in the contrast between Patterns B and C, we also ran these models with Pattern B as the baseline. The random structures in these models were identical.

Results

Scanpath clusters

We identified three reading patterns in both datasets. Results from the linear models showing differences in reading time measures across the three clusters are shown in Table 3, and the reading patterns are illustrated in Figure 2. Pattern A was primarily linear; in other words there were little to no regressions back to previous parts of the texts after the reader reached the target phrase. We labeled it as *fast-and-linear reading* pattern. Pattern B consisted of more re-reading, which seems to be targeted to the context, and around the target sentence, and we call this *selective rereading* pattern. Pattern C had the longest reading times with extensive re-readings of the whole text, indicating that it reflects *nonselective rereading*. Overall, the fast-and-linear reading pattern was the most common in the data (61% of scanpaths), followed by the selective rereading pattern (29% of scanpaths), and the non-selective rereading pattern was used the least (10% of scanpaths).

Scanpaths associated with text type

We first ran models with only text type as predictor (Model 1) to investigate whether cluster membership could be predicted by whether participants were reading the ironic or literal texts. There was a main effect of text type in the comparison between fast-and-linear reading and non-selective rereading patterns, as well as between selective and nonselective rereading patterns. These effects indicate that nonselective rereading was the least likely pattern to be used, and that this difference was even larger with ironic texts.

WMC and scanpaths

We then ran models (Model 2) investigating the effect of working memory, text type, and their interaction on the likelihood of demonstrating different reading patterns (see Tables 4 and 5). There were main effects of text type such that nonselective rereading pattern was less likely to be used for ironic texts compared to fast-and-linear reading and selective rereading patterns. There was also a main effect of working memory such that higher WMC was associated with a lower probability of using selective compared to nonselective rereading pattern. There was also a significant interaction between text type

Table 3. Model estimates of reading times per cluster.

Measure	Cluster		
	Pattern A	Pattern B	Pattern C
Target Sentence			
First-Pass Reading	820 (40)	864 (43)	860 (46)
First-Pass Re-Reads	280 (12)	274 (13)	302 (20)
Looks FROM	376 (16)	513 (24)	980 (62)
Looks TO	376 (17)	545 (27)	955 (61)
Spillover Region			
First-Pass Reading	704 (65)	750 (70)	764 (77)
First-Pass Re-Reads	309 (22)	356 (28)	439 (45)
Critical Context			
Looks TO	390 (24)	672 (42)	1633 (121)

Note. Table 3 shows model estimated reading times and standard error on areas of interest for each cluster on the target sentence, the spillover region, and the critical context. Pattern A = fast-and-linear reading pattern, Pattern B = selective rereading pattern, and Pattern C = nonselective rereading pattern.

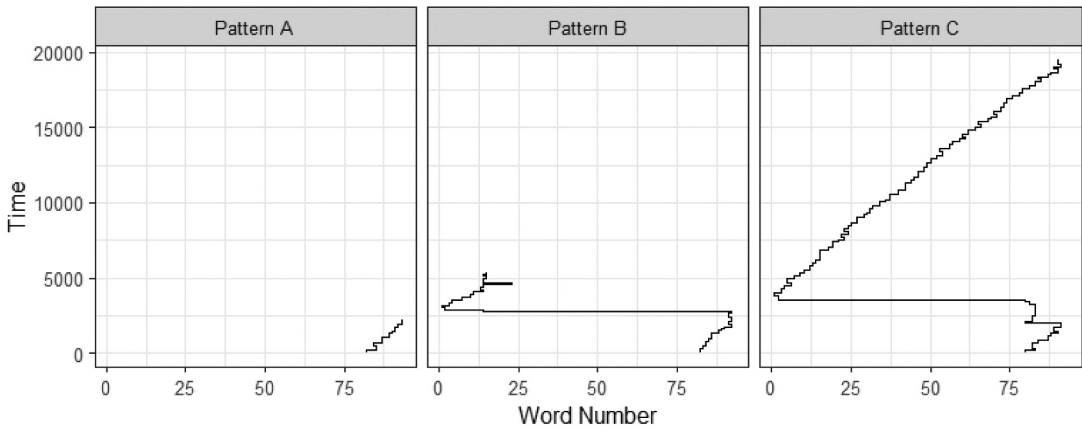


Figure 2. Examples of the three scanpaths identified in the datasets. *Note.* Figure 2 shows example scanpaths for the three patterns identified by the cluster analysis. Pattern A = fast-and-linear reading pattern, Pattern B = selective rereading pattern, and Pattern C = nonselective rereading pattern.

Table 4. Model estimates of the cluster membership with Pattern A as the baseline.

	Model 1		Model 2		Model 3		Model 4	
	Pattern B	Pattern C	Pattern B	Pattern C	Pattern B	Pattern C	Pattern B	Pattern C
<i>Intercept</i>	-0.74*	-2.20*	-0.78*	-2.26*	-0.76*	-2.23*	-0.77*	-2.24*
	(0.15)	(0.21)	(0.15)	(0.22)	(0.15)	(0.21)	(0.15)	(0.21)
<i>Text Type (Literal vs. Ironic)</i>	-0.05	-0.42*	-0.02	-0.46*	-0.04	-0.44*	-0.04	-0.43*
	(0.11)	(0.18)	(0.12)	(0.18)	(0.11)	(0.18)	(0.11)	(0.18)
<i>WMC</i>			-0.26	0.11				
			(0.13)	(0.21)				
<i>Text Type x WMC</i>			0.10	-0.34				
			(0.13)	(0.20)				
<i>Memory Questions</i>					-0.05	0.18		
					(0.12)	(0.20)		
<i>Text type x Memory Questions</i>					0.16	0.29		
					(0.11)	(0.18)		
<i>Inference Questions</i>							0.06	0.17
							(0.12)	(0.21)
<i>Text Type x Inference Questions</i>							0.15	0.11
							(0.11)	(0.19)

Note. Table 4 shows the output of the multinomial models, with Pattern A as the baseline. All estimates are in log odds. * = $p < .05$. Pattern A = fast-and-linear reading pattern, Pattern B = selective rereading pattern, and Pattern C = nonselective rereading pattern.

and working memory in the comparison between selective rereading and nonselective rereading patterns (see Figure 3). This interaction indicates that when the readers' WMC increased, they were more likely to use nonselective in comparison to selective rereading pattern for literal texts, whereas for ironic texts WMC did not affect the likelihood of using nonselective rereading pattern, but selective rereading pattern was more likely to be used. There were no significant differences between fast-and-linear reading and selective rereading patterns, although there was a trend for a main effect of WMC such that as WMC increased participants were more likely to use the fast-and-linear reading pattern.

Associations between comprehension and scanpaths

Last, we ran models (Models 3 and 4) investigating the relationship between reading patterns, text type (literal vs. ironic), and comprehension accuracy (text memory and inferential questions as separate measures). There were no significant effects of comprehension accuracy nor any interactions.

Table 5. Output of multinomial models with Pattern B as baseline.

	Model 1	Model 2	Model 3	Model 4
	Pattern C	Pattern C	Pattern C	Pattern C
Intercept	-1.52* (0.17)	-1.52* (0.16)	-1.53* (0.16)	-1.52* (0.16)
Text Type (Literal vs. Ironic)	-0.37* (0.17)	-0.39* (0.18)	-0.37* (0.18)	-0.40* (0.18)
WMC		0.35* (0.16)		
Text Type x WMC		-0.43* (0.20)		
Memory Questions			0.24 (0.15)	
Text Type x Memory Questions			0.12 (0.17)	
Inference Questions				0.11 (0.16)
Text Type x Inference Questions				-0.05 (0.18)

Note. Table 5 shows the output of the multinomial models with Pattern B as the baseline compared to Pattern C. All estimates are in log odds. * = $p < .05$. Comparison with Pattern A is shown in Table 4, hence, this comparison is not repeated here. Pattern A = fast-and-linear reading pattern, Pattern B = selective rereading pattern, and Pattern C = nonselective rereading pattern.

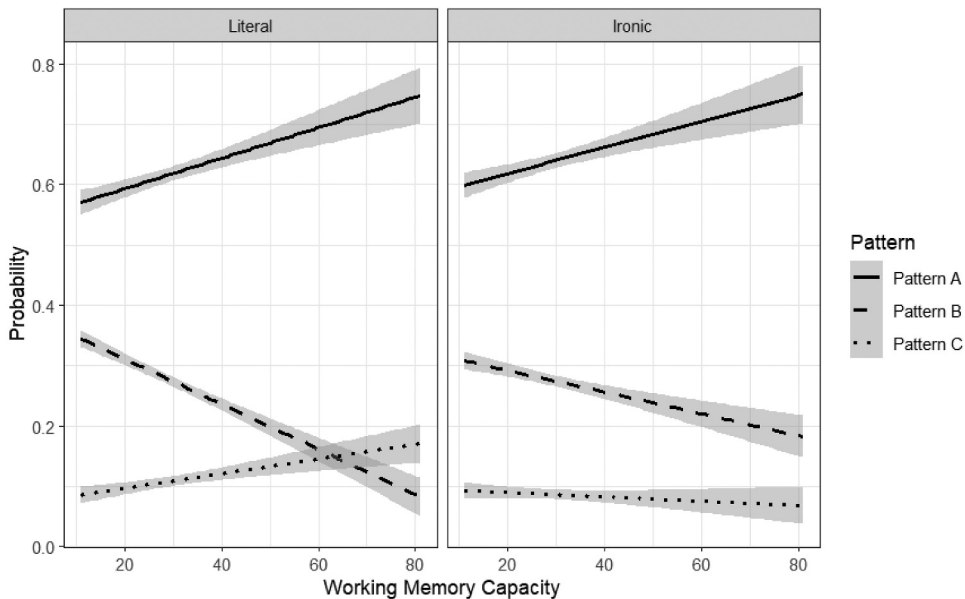


Figure 3. Model estimated probabilities of different scanpath Patterns (A, B, and C) as a function of WMC score separately for literal and ironic texts. Note. The gray areas indicate 95% CI. Pattern A = fast-and-linear reading pattern, Pattern B = selective rereading pattern, and Pattern C = nonselective rereading pattern.

Discussion

In the present study, for the first time, we used scanpath analysis to investigate processing of written irony. Moreover, we studied how WMC and comprehension of irony are reflected in the readers' scanpaths. The cluster analysis revealed that readers used three different scanpath patterns: First, in the fast-and-linear reading pattern readers showed none or very little rereading. Second, in the selective rereading pattern readers made more re-reading, which seemed to be strategic. In other words, rereading was selectively targeted to the context, and around the target sentence. Last, in the nonselective rereading pattern readers showed the longest reading times with extensive nonselective re-reading of the whole text. Interestingly, these scanpath patterns resemble reading strategies previously identified for reading expository texts (e.g., Hyönä & Nurminen, 2006; Hyönä et al., 2002; León et al., 2019).

Overall, readers were less likely to use nonselective rereading pattern than fast-and-linear reading or selective rereading patterns when reading ironic compared to literal texts. In other words, it seems that reading ironic texts is pretty straightforward at the whole text-level, and when reprocessing is needed, it is selective and directed toward interpretation-relevant text regions (i.e., target phrase and context). With the literal texts there is no discrepancy between the context and the target phrase, which would trigger selective reprocessing of these text regions, and reading is more nonselective. This notion is supported by more general reading-related findings by Hyönä et al. (2002), who showed that the fast-and-linear reading pattern was the most often used while reading expository texts. Similarly to irony, in expository texts arguments are backed up with contextual information, which seems to guide reading behavior and encourage rereading of the interpretation-relevant regions of text.

The finding that fast-and-linear and selective rereading reading patterns were the most used with irony is in line with previous eye-tracking studies using word- and sentence-level measures. These studies have shown that processing written irony is characterized by rereading of the target phrase and interpretation-relevant parts of context (see Olkonemi & Kaakinen, 2021 for review). This said, it should be noted that the nonselective rereading pattern was generally the least likely to be used (overall 10% of the scanpaths) and the fast-and-linear reading pattern was the most used with the literal texts as well. It seems that at the global level of text processing, irony is often processed similarly to literal text. This result partly supports the *direct access view* suggesting that people could process ironic language as effortlessly as literal language (e.g., Gibbs, 1994; Gibbs & O'Brien, 1991). From the methodological point of view, this means that the scanpath analysis reflects processing differences at a different level than the traditional eye movement measures reported in previous studies (e.g., Filik & Moxey, 2010; Kaakinen et al., 2014). Whereas scanpath analysis reflects global strategies in how reader's eye gaze transfers between different parts of text, traditional measures offer information about how much time a reader spends on specific parts of text during first-pass and subsequent rereadings.

The reading patterns of ironic and literal texts were modulated by readers' WMC. Overall, high WMC readers were overall more likely to use the fast-and-linear reading pattern than low WMC readers. In addition, high WMC readers preferred to use the selective over nonselective rereading pattern for irony, whereas the opposite was true for literal texts. For readers with relatively low WMC, in contrast, reading patterns between literal and ironic texts did not differ. This suggests, first, that readers with relatively high WMC tend to engage in more linear reading than those with low WMC. This is in line with previous studies, which have shown that high WMC readers are able to start processing ironic meaning as soon as they encounter irony in text (e.g., "What great weather for a picnic!" when it's raining heavily; see Table 1), whereas low WMC readers need to rely more on later rereading (Kaakinen et al., 2014; Olkonemi et al., 2016, see also Columbus et al., 2015). This further suggests that higher WMC would allow readers to maintain prior text information in their memory and then adopt the interpretation that proves to be consistent with the context (e.g., Just & Carpenter, 1992).

Second, readers with high WMC are more sensitive to the presence of irony as reflected in differences between ironic and literal texts in the scanpath patterns, whereas readers with lower WMC show similar scanpath patterns for literal and ironic texts. It seems like readers with higher WMC would be more flexible in changing from a more selective rereading pattern with ironic texts to a nonselective rereading pattern with literal texts, whereas readers with lower WMC use similar reading patterns with both kinds of texts. This interpretation is in line with the idea that high WMC readers are better in adapting to the task demands and maintaining attentional control when the task demands it (see Engle, 2018). This is also partly in line with studies showing that high WMC readers are likely to use a more strategic reading pattern while reading expository texts (Hyönä et al., 2002).

Findings related to the relationship between WMC and scanpath patterns during reading ironic and literal texts support newer theories on irony comprehension that explicitly take into account individual differences (Fabry, 2021; Pexman, 2008). For example, *predictive coding theory* on irony (Fabry, 2021) assumes that individuals with higher WMC would be better in making predictions

during reading, because they are better at maintaining relevant priors in an accessible state for the subsequent interpretation process (Trapp et al., 2021). They may also be better at recovering from prediction errors. Here the results suggest that high WMC increases the use of a selective rereading pattern during reading of ironic texts. It is hard to say whether this reflects better prediction accuracy or faster recovery from prediction errors, as the materials used in the present study (Kaakinen et al., 2014; Olkonieni et al., 2016) did not contain texts in which irony could have been predicted from the context. Thus, further studies are needed to explore which theoretical views can account for individual differences in irony processing.

Last, we failed to find a relationship between scanpath patterns and irony comprehension. Previous eye-tracking studies showing the relationship between irony processing and comprehension have either studied children, who show clearly lower comprehension scores (Olkonieni et al., 2023), or they have been experiments using a masking paradigm disrupting normal reading (Olkonieni, Johander, et al., 2019). Moreover, previous studies have reported reading times on specific areas of interest within the text, and not looked at global scanpath patterns. One possible reason for the failure to observe a relationship between scanpaths and comprehension is that there was limited variance in comprehension scores in the present datasets – the mean irony comprehension accuracy was $\cong 83\%$. This suggests that although previous studies have shown that irony comprehension often is difficult, adult readers are good at making a successful interpretation when encountering ironic phrases. Further studies are needed to explore the relationship between processing and comprehension of irony.

Conclusions

The present study used scanpath analysis for the first time to study processing and comprehension of written irony. We identified three reading patterns: fast-and-linear reading, selective rereading, and nonselective rereading patterns. The results showed that encountering irony in text guided reading toward more fast-and-linear reading and selective rereading patterns over nonselective rereading patterns. Moreover, the results showed that this effect was modulated by WMC. It seems that higher WMC readers are more flexible in changing their reading pattern based on the task demands by changing from being more likely to use nonselective rereading patterns with literal texts to using more selective rereading with ironic texts. Overall, the present study also shows that the scanpath methodology is a valuable new tool for analyzing eye-movement data on figurative language processing. However, as it captures a different level of eye movement behavior than the traditional eye movement measures, it is not likely to completely replace the traditional measures.

Acknowledgments

This study was supported by Academy of Finland grant #338712 awarded to Henri Olkonieni and Academy of Finland project grant #334266 awarded to Johanna K. Kaakinen. Parts of this paper were presented at the 33rd Annual Meeting of the Society for Text and Discourse. We would like to thank Dr. Titus von der Malsburg for his advice on the scanpath analysis.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Academy of Finland [338712].

ORCID

Henri Olkonemi  <http://orcid.org/0000-0001-6915-089X>

Diane Mézière  <http://orcid.org/0000-0002-5406-4590>

Johanna K. Kaakinen  <http://orcid.org/0000-0002-2970-826X>

References

- Antoniou, K., & Milaki, E. (2021). Irony comprehension in bidialectal speakers. *The Modern Language Journal*, 105(3), 697–719. <https://doi.org/10.1111/modl.12724>
- Attardo, S. (2000). Irony as relevant inappropriateness. *Journal of Pragmatics*, 32(6), 793–826. [https://doi.org/10.1016/S0378-2166\(99\)00070-3](https://doi.org/10.1016/S0378-2166(99)00070-3)
- Au-Yeung, S. K., Kaakinen, J. K., Liversedge, S. P., & Benson, V. (2015). Processing of written irony in autism spectrum disorder: An eye movement study. *Autism Research*, 8(6), 749–760. <https://doi.org/10.1002/aur.1490>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1). <https://doi.org/10.18637/jss.v067.i01>
- Columbus, G., Sheikh, N. A., Côté-Lecaldare, M., Häuser, K., Baum, S. R., & Titone, D. (2015). Individual differences in executive control relate to metaphor processing: An eye movement study of sentence reading. *Frontiers in Human Neuroscience*, 8. <https://doi.org/10.3389/fnhum.2014.01057>
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19(4), 450–466. [https://doi.org/10.1016/S0022-5371\(80\)90312-6](https://doi.org/10.1016/S0022-5371(80)90312-6)
- Elff, M. (2022). *Mclogit: Multinomial logit models, with or without random effects or overdispersion* (Version 0.9.6). <https://CRAN.R-project.org/package=mclogit>
- Engle, R. W. (2018). Working memory and executive attention: A revisit. *Perspectives on Psychological Science*, 13(2), 190–193. <https://doi.org/10.1177/1745691617720478>
- Fabry, R. E. (2021). Getting it: A predictive processing approach to irony comprehension. *Synthese*, 198(7), 6455–6489. <https://doi.org/10.1007/s11229-019-02470-9>
- Ferreira, F., & Yang, Z. (2019). The problem of comprehension in psycholinguistics. *Discourse Processes*, 56(7), 485–495. <https://doi.org/10.1080/0163853X.2019.1591885>
- Filik, R., Howman, H., Ralph-Nearman, C., & Giora, R. (2018). The role of defaultness and personality factors in sarcasm interpretation: Evidence from eye-tracking during reading. *Metaphor and Symbol*, 33(3), 148–162. <https://doi.org/10.1080/10926488.2018.1481258>
- Filik, R., & Moxey, L. M. (2010). The on-line processing of written irony. *Cognition*, 116(3), 421–436. <https://doi.org/10.1016/j.cognition.2010.06.005>
- Friston, K., Kilner, J., & Harrison, L. A. (2006). A free energy principle for the brain. *Journal of Physiology – Paris*, 100(1–3), 70–87. <https://doi.org/10.1016/j.jphysparis.2006.10.001>
- Gibbs, R. W. (1994). *The poetics of mind: Figurative thought, language, and understanding*. Cambridge University Press.
- Gibbs, R. W., & O'Brien, J. (1991). Psychological aspects of irony understanding. *Journal of Pragmatics*, 16(6), 523–530. [https://doi.org/10.1016/0378-2166\(91\)90101-3](https://doi.org/10.1016/0378-2166(91)90101-3)
- Giora, R. (1999). On the priority of salient meanings: Studies of literal and figurative language. *Journal of Pragmatics*, 31(7), 919–929. [https://doi.org/10.1016/S0378-2166\(98\)00100-3](https://doi.org/10.1016/S0378-2166(98)00100-3)
- Giora, R. (2003). *On our mind: Salience, context, and figurative language*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195136166.001.0001>
- Godbee, K., & Porter, M. (2013). Comprehension of sarcasm, metaphor and simile in Williams syndrome. *International Journal of Language & Communication Disorders*, 48(6), 651–665. <https://doi.org/10.1111/1460-6984.12037>
- Green, J. (2012). *The fault in our stars*. Penguin Group.
- Grice, H. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics 3: Speech acts* (pp. 41–58). Academic Press.
- Hyönä, J., Lorch, R. F., Jr, & Kaakinen, J. K. (2002). Individual differences in reading to summarize expository text: Evidence from eye fixation patterns. *Journal of Educational Psychology*, 94(1), 44–55. <https://doi.org/10.1037/0022-0663.94.1.44>
- Hyönä, J., & Nurminen, A.-M. (2006). Do adult readers know how they read? Evidence from eye movement patterns and verbal reports. *British Journal of Psychology*, 97(1), 31–50. <https://doi.org/10.1348/000712605X53678>
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99(1), 122–149. <https://doi.org/10.1037/0033-295X.99.1.122>
- Kaakinen, J. K. (2017). On-line measures of text processing. In M. F. Schober, D. N. Rapp, & M. A. Britt (Eds.), *The Routledge handbook of discourse processes* (2nd ed., pp. 125–130). Routledge. <https://doi.org/10.4324/9781315687384-8>
- Kaakinen, J. K., & Hyönä, J. (2007). Strategy use in the reading span test: An analysis of eye movements and reported encoding strategies. *Memory*, 15(6), 634–646. <https://doi.org/10.1080/09658210701457096>

- Kaakinen, J. K., Olkonemi, H., Kinnari, T., & Hyönä, J. (2014). Processing of written irony: An eye movement study. *Discourse Processes*, 51(4), 287–311. <https://doi.org/10.1080/0163853X.2013.870024>
- Kreuz, R. J., & Link, K. E. (2002). Asymmetries in the use of verbal irony. *Journal of Language and Social Psychology*, 21(2), 127–143. <https://doi.org/10.1177/02627X02021002002>
- León, J. A., Moreno, J. D., Escudero, I., & Kaakinen, J. K. (2019). Selective attention to question-relevant text information precedes high-quality summaries: Evidence from eye movements. *Journal of Eye Movement Research*, 12(1). <https://doi.org/10.16910/jemr.12.1.6>
- McCabe, D. P., Roediger, H. L., McDaniel, M. A., Balota, D. A., & Hambrick, D. Z. (2010). The relationship between working memory capacity and executive functioning: Evidence for a common executive attention construct. *Neuropsychology*, 24(2), 222–243. <https://doi.org/10.1037/a0017619>
- Mézière, D., Yu, L., McArthur, G., Reichle, E., & von der Malsburg, T. (2022). Scanpath regularity as an index of reading comprehension. *PsyArXiv*. <https://doi.org/10.31234/osf.io/w6x4t>
- Olkonemi, H., Halonen, S., Pexman, P. M., & Häikiö, T. (2023). Children's processing of written irony: An eye movement study. *Cognition*, 238, 105508. <https://doi.org/10.1016/j.cognition.2023.105508>
- Olkonemi, H., Johander, E., & Kaakinen, J. K. (2019). The role of look-backs in the processing of written sarcasm. *Memory & Cognition*, 47(1), 87–105. <https://doi.org/10.3758/s13421-018-0852-2>
- Olkonemi, H., & Kaakinen, J. K. (2021). Processing of irony in text: A systematic review of eye-tracking studies. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, 75(2), 99–106. <https://doi.org/10.1037/cep0000216>
- Olkonemi, H., Ranta, H., & Kaakinen, J. K. (2016). Individual differences in the processing of written sarcasm and metaphor: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(3), 433–450. <https://doi.org/10.1037/xlm0000176>
- Olkonemi, H., Strömberg, V., & Kaakinen, J. K. (2019). The ability to recognise emotions predicts the time-course of sarcasm processing: Evidence from eye movements. *The Quarterly Journal of Experimental Psychology*, 72(5), 1212–1223. <https://doi.org/10.1177/1747021818807864>
- Parshina, O., Sekerina, I. A., Lopukhina, A., & von der Malsburg, T. (2022). Monolingual and bilingual reading processes in Russian: An exploratory scanpath analysis. *Reading Research Quarterly*, 57(2), 469–492. <https://doi.org/10.1002/rrq.414>
- Pexman, P. M. (2008). It's fascinating research: The cognition of verbal irony. *Current Directions in Psychological Science*, 17(4), 286–290. <https://doi.org/10.1111/j.1467-8721.2008.00591.x>
- R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <http://www.R-project.org/>
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372–422. <https://doi.org/10.1037/0033-2909.124.3.372>
- Rayner, K. (2009). The 35th sir frederick bartlett lecture: Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, 62(8), 1457–1506. <https://doi.org/10.1080/17470210902816461>
- Scrucca, L., Fop, M., Murphy, T. B., & Raftery, A. E. (2016). mclust 5: Clustering, classification and density estimation using Gaussian finite mixture models. *The R Journal*, 8(1), 289–317. <https://doi.org/10.32614/RJ-2016-021>
- Spotorno, N., & Noveck, I. (2019). Ironic utterances. In C. Cummings & N. Katsos (Eds.), *The Oxford handbook of experimental semantics and pragmatics* (pp. 283–297). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198791768.013.16>
- Trapp, S., Parr, T., Friston, K., & Schröger, E. (2021). The predictive brain must have a limitation in short-term memory capacity. *Current Directions in Psychological Science*, 30(5), 384–390. <https://doi.org/10.1177/09637214211029977>
- Turcan, A., & Filik, R. (2016). An eye-tracking investigation of written sarcasm comprehension: The roles of familiarity and context. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(12), 1867–1893. <https://doi.org/10.1037/xlm0000285>
- Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S* (4th ed.). Springer.
- von der Malsburg, T. (2018). *Scanpath: Tools for analyzing spatio-temporal patterns in eye movements* (Version 1.06). <https://github.com/tmalsburg/scanpath>
- von der Malsburg, T., & Vasishth, S. (2011). What is the scanpath signature of syntactic reanalysis? *Journal of Memory and Language*, 65(2), 109–127. <https://doi.org/10.1016/j.jml.2011.02.004>
- von der Malsburg, T., & Vasishth, S. (2013). Scanpaths reveal syntactic underspecification and reanalysis strategies. *Language and Cognitive Processes*, 28(10), 1545–1578. <https://doi.org/10.1080/01690965.2012.728232>
- Walczyk, J. J., & Taylor, R. W. (1996). How do the efficiencies of reading subcomponents relate to looking back in text? *Journal of Educational Psychology*, 88(3), 537–545. <https://doi.org/10.1037/0022-0663.88.3.537>