

Analysing revisions in online writing

Abstract

New technology has given new insights into the writing process of professional text producers. Keystroke logging can give the researcher knowledge about how text producers write and revise their texts, both at the level of form and content. With a point of departure in the results of Haugaard's (2016) study of journalistic writing, this article suggests a reassessment of the online revision taxonomy developed by Lindgren/Sullivan (2006b), with a focus on external revisions. In order to be able to interpret the effect of online revisions on the text by means of keystroke logging, it is proposed that revisions be analysed with a point of departure in the semantic content involved, rather than according to location, i.e. in the text already transcribed (*contextual revision*) or in the text currently being transcribed (*pre-contextual revision*), as suggested by Lindgren/Sullivan (2006b). It is argued that contextual and pre-contextual revisions should not be conceptualised as dichotomous entities, but as open categories on a continuum of semantically meaningful context, on the basis of which revisions can be interpreted depending on the degree of completeness of the context in which they are made.

Keywords

writing; revision of form and content; professional text production; online revision taxonomy; keystroke logging; continuum

1. Introduction

Writing is an intricate process which requires cognitive resources, for example motivation, knowledge, memory, and creativity, as well as interaction with situational elements, such as internet searches, use of various handbooks, conferring with source material, experts and colleagues, and adaptation to corporate standards. All these different instances of the writing process feed into and shape the way the text is written, influencing the use of strategies with regard to planning, generating, formulating, evaluating, and revising and vice versa (Van Waes/Schellens 2003: 830). Revision can both refer to the cognitive process of changing a text and to the changes themselves (e.g. Fitzgerald 1987: 483, Freedman 1985: xi and Alamargot/Chanquoy 2001: 100). In this article we zoom in on revision as it unfolds in online text production as changes made at the surface of the text of form and content.

The article reports on and develops Haugaard's (2016) method for analysing online revision of form and content with Rikke Hartmann Haugaard as the first author and Helle Dam-Jensen as second author. On the basis of insights made by for example Perl (1979), Bridwell (1980), Sommers (1980), Kaufer et al. (1986), Matsuhashi (1987), and Chenoweth/Hayes (2001, 2003), Haugaard (2016) assumes that revisions are not only carried out as revisions in a completed draft or text, but also as revisions carried out throughout the text-production process. Building on Lindgren/Sullivan's (2006a and 2006b) taxonomy of revision in online text production, Haugaard (2016) investigates revisions of form and content by means of keystroke logging, combined with retrospective interviews for disambiguating unclear log files.

* Rikke Hartmann Haugaard
Department of Culture and Learning
Aalborg University
E-mail: haugaard@cgs.aau.dk

Helle Dam Jensen
Spanish Business Communication
Aarhus University
E-mail: hed@cc.au.dk

Before presenting the method for analysing online revisions, the next section sketches revision and its location in the writing process.

2. The cognitive writing process

The first studies into writing described the writing process as a linear activity, in which writers first planned what they were going to write, then they wrote a draft, which they then revised (Hill et al. 1991). In this conception, revision is “what the writer does after a draft is completed” (Murray 1978: 87), a statement which was also reflected in the concepts used by Rohman/Wlecke (1964): ‘Pre-writing’, ‘Writing’ and ‘Re-writing’, and Murray (1978): ‘Prevision’, ‘Vision’ and ‘Revision’. However, investigation showed that the writing process is of a much more complex nature than a linear presentation made it out to be. Thus, a new understanding of writing evolved which explained the process as a hierarchically structured cognitive process, in which writing is carried out in recursive loops (Flower/Hayes 1981: 367ff., Nystrand 2001: 115-117, Krogh 2003: 242). The cognitive process includes various elements, which writers orchestrate during writing “to progressively create a text that meets their conception of topic, task and audience” (Lindgren/ Sullivan 2006a: 32). Studies on writing vary in the degree of elaborateness with which the process is spelled out and in the way the elements are named. Nevertheless, common elements seem to be planning, generating, formulating, evaluating, and revising. The recursivity and the hierarchical structure of the process imply that revision “[...] can interrupt other composing processes at any given time [...]” (Witte 1985: 259) and, thus, “occur[s] continually throughout the writing of a work” (Sommers 1980: 380). Accordingly, revision is not (only) “an end-of-the-line repair process” (Flower/Hayes 1981: 367) but is, in fact, an inherent part of text production. In this sense, and according to Nold (1979: 105-106 in Fitzgerald 1987: 483), revision does not only amount to the correction of downright errors, but also applies to, for example, the changing of the content of the text, improvements in the formulation of content, and changes in the order of text segments.

3. Revision

Various studies have investigated revisions of the written text. Hildick (1965), Stallard (1974) and NAEP (1977) developed some of the earliest classifications, thus pioneering the investigation of revision. By way of example, Stallard (1974) developed a classification of six revision types on the basis of an inductively conducted study of 30 high school students, divided into a group of 15 “good writers” and a comparison group composed of 15 randomly selected students. The methodological tools used were observation, interview and analysis of handwritten material from the whole writing process. It goes without saying that Stallard (1974), like other researchers at the time, was subject to constraints in terms of the investigation methods available, which restricted the insights into the process that could be gained. Fitzgerald (1987: 485) summarises other critical points directed at Stallard (1974), as well as at Hildick (1965) and NAEP (1977). She points out that revision categories were not mutually exclusive, which resulted in overlaps, and that important revision types, such as addition and omission of text, were not established as taxonomic categories.

These taxonomic problems were addressed by, among others, Bridwell (1980) and Sommers (1980). Their studies revealed more revision types than those depicted in previous classifications, for example, addition, omission, and substitution, which resulted in more comprehensive classifications. The fact that the classifications also specified the syntactic level at which a revision operates, e.g. word, phrase, or sentence, contributed to a more accurate representation of the intricacies of revising. Additionally, a clear delineation of revision types and syntactic levels overcome the problem of previous classification with categories that were not mutually exclusive (Fitzgerald 1987: 486). The drawback of both Bridwell and Sommers’ (1980) classifications is that they were restricted to classification of revisions at specific points of the process, as either

in-process revisions (revisions made during the writing of a draft) or as revisions carried out between drafts (revisions made in a completed first draft to be applied in a second revised draft) (Fitzgerald 1987: 486). When it comes to the investigation of contemporary text production, such classifications give a distorted picture of the writing process. Furthermore, by tying attention to specific points of the process in detriment of others, they can be said to impede investigation of all parts of the process.

This gap is remedied by Matsushashi (1987), who developed a notational framework for tracking revisions on the basis of a study of a student's revisions and pauses in handwritten text production. The method applied was (timed) video recordings allowing for the registration of the successive revisions made during writing and their sequence. Matsushashi (1987) found that, in relation to the leading edge of the text, revisions were carried out in three different contexts. 'Sentence-immediate revisions' are carried out in the word or 'unit of language' (known as 'bursts' in later studies) that has just been transcribed, 'sentence-distal revisions' are made elsewhere in the current sentence, and 'text revisions' are carried out elsewhere in the text (Matsushashi 1987: 212).

Matsushashi's (1987) methodological approach was innovative and has inspired later studies of revision. As an example, Severinson Eklundh/Kollberg (1996) developed a computerised notation, the S-notation, on the basis of Matsushashi's (1987) seminal work. Although Matsushashi (1987) as well as Severinson Eklundh/Kollberg (1996) provide a fine-grained and sophisticated analytical framework, they do not provide the analytical means for studying revision of form and content. Fitzgerald (1987: 486) notes that the first study that discriminated between revision of form and content were Faigley/Witte (1981). Starting from Bridwell's procedures for gathering data, Faigley/Witte (1981) developed a taxonomy of revisions that made a distinction between surface changes (form) and meaning changes (content) as well as between microstructure and macrostructure features. Moreover, six revision types were identified (for example addition and substitution). Their revision taxonomy stands out as one of the most influential models for revision analysis (Lindgren/Sullivan 2006a: 40, Chanquoy 2009: 87). However, in Faigley/Witte's (1981) study, form and content apply to the product, rather than to the process (see Humes' (1983: 204) label 'quasi-product study'). On a general level, this means that revisions in product studies will be analysed and interpreted in relation to the entire text, for example by comparing different drafts, and not in relation to the text transcribed at the time of revision, which is the case of revisions in process studies (see also Lindgren/Sullivan 2006b: 180).

With the advent of the computer, the text-production process changed in significant ways, thus bringing new methodological aids to the researcher. Various revision studies have been carried out during the last decades, e.g. Kollberg/Severinson Eklundh (2001), Perrin (2003, 2013), Van Waes/Schellens (2003), Lindgren/Sullivan (2006b), Stevenson et al. (2006), and Van Hout (2010), all of them applying keystroke logging. In the interest of this article, Lindgren/Sullivan (2006b) stand out in that they offer a framework that takes form and content into account. Building on insights of form and content (Faigley/Witte 1981), process (Matsushashi 1987), and revision categories (Bridwell 1980 and Sommers 1980), Lindgren/Sullivan develop a taxonomy of online revisions, as we will see in the next section.

4. The LS-taxonomy

Observation of the text-production process as it unfolds in computer-based writing reveals the continuous revisions made by the writer. At any given point during writing, the written text can either be revised at its leading edge¹, where new text is being transcribed, or it can be revised in the text already written, i.e. after the text has been transcribed. These online revisions contribute

1. We suggest the term leading edge, coined by Baaijen et al. (2012) rather than 'point of inscription' used by many researchers (e.g. Matsushashi 1987 and Stevenson/Schoonen/de Gloppe 2006) because it allows us to distinguish between the current end of the text and any other position of the cursor where text is inscribed.

to develop the emergent text imposing specific requirements on the analytical and interpretative work of the researcher.

We saw above that different revision taxonomies have been developed, which can be used as a methodological tool for analysing revisions in text production. One such taxonomy is the LS-taxonomy (Figure 1), developed by Lindgren and Sullivan (2006a and 2006b) in collaboration with Stevenson (Stevenson et al. 2006), on the basis of an empirical study of young Swedish text producers (13-15 years) composing texts in both Swedish and English. The taxonomy aims at capturing the specific part of the composition process that involves changes in form and concepts (content). To this end, an overall distinction between internal and external revisions is made, depending on whether revisions are made before transcribing (internal) or after transcribing (external), thus recognising that revision occurs at the mental level as well as at the textual level of the text.

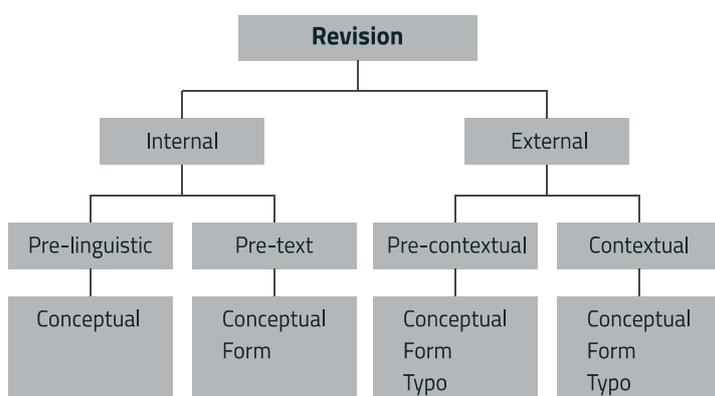


Figure 1: The LS-taxonomy (Lindgren/Sullivan 2006a: 39)

Both internal and external revisions divide into subcategories. Internal revisions are either pre-linguistic or pre-textual, depending on whether the revision affects the conceptual and/or linguistic level of a mental representation or a mental text. Accordingly, pre-linguistic revisions occur before an idea is formulated linguistically affecting only the conceptual level of the text, e.g. a change in a plan or an idea (Lindgren/Sullivan 2006a: 38, Stevenson et al. 2006: 206). Pre-textual revisions, in contrast, are changes in mental text prior to transcription (Lindgren/Sullivan 2006a: 38, Stevenson et al. 2006: 206), a pre-text being “a “trial locution” that is produced in the mind, stored in the writer’s memory, and sometimes manipulated mentally prior to being transcribed as written text.” (Witte 1987: 397). Hence, pre-textual revisions differ from pre-linguistic revisions because they may imply changes to mental representations of both concept and form, although they may be limited to the latter.

As made explicit in the introduction, this paper is restricted to the investigation of external revisions, viz. revisions that are visible in the keystroke-logging files, which will be explained in detail in the following paragraph.

4.1. External revision

As noted above, external revisions are either pre-contextual or contextual depending on whether they are made in the text that has already been transcribed or is currently being transcribed. Lindgren/Sullivan (2006b: 159) define pre-contextual revisions, a term coined by Lindgren (2005:

17), as “those revisions made before an externalized context is completed²”. They characterise the distinguishing features of pre-contextual revisions as follows:

(1) that at the time a pre-contextual revision occurs the only externalised text is before the place of revision, (2) that at the time of the revision there is no externalised text following the place of revision and (3) that at the time of the revision, the revised text represents the writer’s last externalised text item. (Lindgren/Sullivan 2006b: 159)

Hence, pre-contextual revisions are changes in the text just written at the leading edge, i.e. no textual context is transcribed after the revision is carried out, only before.

At the leading edge, the text producer continually makes decisions about the emergent text, that is, what to write and how to formulate the ideas linguistically (Matsushashi 1987: 204, Lindgren 2005: 32) and revisions are made when what has just been transcribed or is being transcribed needs adjustment (Lindgren/Sullivan 2006b: 161). According to Lindgren/Sullivan (2006b: 171), contextual revisions are:

revisions undertaken when writers move away from the point of inscription to insert new text or to delete, substitute or rearrange already written text. When a contextual revision is undertaken, writers are operating within an externalised context; a contextual revision is conducted within a previously written and completed sentence. Hence, a contextual revision is both preceded and followed by text.

As opposed to pre-contextual revisions, contextual revisions are, thus, undertaken when writers move away from the leading edge to insert new text or to exclude, substitute or rearrange already transcribed text. Lindgren/Sullivan (2006b: 171) explicitly state that externalised context is defined as a ‘completed sentence’ and that the revision, consequently, is ‘preceded and followed by text’. The distinction between the two types of external revision can be graphically depicted in Figure 2:

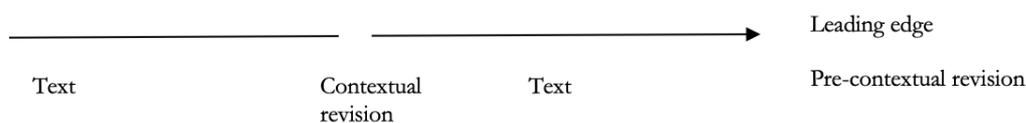


Figure 2: External revision

The distinction between pre-contextual and contextual revisions is relevant when the effect of a revision on the content or the form of the text is to be interpreted. At the leading edge of the text, the future text has not yet been transcribed, which makes it impossible to know what the writer intends to write after the revision has been made (Lindgren/Sullivan 2006a: 43). Therefore, the effect of pre-contextual revisions is interpreted based on the preceding text, which is the only part of the process that the researcher can safely assume the writer to be conscious of when making the revision. Accordingly, pre-contextual revisions lack the analytical context, which characterises contextual revisions, making their effect on the text difficult – even impossible – to interpret only on the basis of an observation of the writing process, e.g. by keystroke logging.

It appears from Figure 1 above that both pre-contextual and contextual revisions may involve changes of concepts and of form (form, typo). Lindgren/Sullivan (2006b) rely on Faigley/Witte’s (1981) interpretation of the effect of revisions of the text. While Lindgren/Sullivan use the term conceptual, Faigley/Witte (1981) use the term meaning, which does not exclusively refer to concepts³ explicitly referred to by the text, but also to concepts which reasonably can be inferred from it. Thus, revisions of meaning either bring new information to the text which cannot be inferred or remove old information “in such a way that it cannot be recovered through drawing inferences” (Faigley/Witte 1981: 402). In the discussion below and in the reassessment of the LS-taxonomy, we will use the term ‘content’, which refers univocally to the semantics of words.

In sum, Lindgren/Sullivan's taxonomy is a process-oriented, online approach to the study of external revisions which categorises revisions according to where they are made in relation to the text that has already been transcribed or is currently being transcribed (contextual or pre-contextual) and according to their effect on the text written thus far, i.e. whether they affect the content or the form (conceptual, form and typo).

4.2. Discussion and reassessment of the LS-Taxonomy

As in any empirical investigation, theoretical statements must lend themselves to testing. When confronted with new data, their explanatory power may be questioned. The LS-taxonomy was tested in an analysis of the text-production process of Spanish journalists by Haugaard (2016) and the results revealed a number of problematic issues related, particularly, to the defining features of contextual revisions, which, subsequently, lead to a reassessment of the taxonomy. The discussion presented in this section concerns the defining criteria for classifying a revision as either contextual or pre-contextual. Our discussion in this section is structured as follows:

contextual revision:

- completed sentence
- preceded and followed by text
- additions at the end of a completed sentence

pre-contextual revision:

- pseudo pre-contextual revisions

4.2.1. Contextual revisions and the feature 'completed sentence'

As appears from the description of Lindgren/Sullivan's (2006b) two external revision categories in the section 4.1 above, they are defined by contrast on the basis of 'completed sentence' which defines the context in which a contextual revision occurs. It is, therefore, problematic that Lindgren/Sullivan fail to offer a definition of 'completed sentence'. Without a proper definition, it is difficult to operationalize revisions carried out "within a previously written and completed sentence" and "before an externalized context is completed" (see section 4.1. above) as two opposing categories in an analysis. The concept cannot be defined on the basis of their empirical examples either, as these are presented in isolation, rarely followed by meta-text explaining where and when a particular example appears in the general composition process (Lindgren/Sullivan 2006b: 174-179). In addition, some examples are presented with a final full stop, while others are presented without, which, therefore, resemble pre-contextual revisions at the leading edge rather than contextual revisions in a completed sentence.

It could be assumed that the authors rely on a common understanding of 'completed sentence', that is, as a sentence ended by a full stop or another sentence-completing character. However, if a defining feature of a contextual revision is that it is carried out in a sentence ended by a full stop, not all revisions lend themselves to categorisation by the LS-taxonomy. Revisions at the level of the phrase, e.g. in a noun phrase, undertaken before the sentence is completed (by a full stop), can, therefore, not be classified as a contextual revision, nor as a pre-contextual revision if it is not carried out at the leading edge of the text. Consequently, such revisions cannot be classified as contextual revisions, nor as pre-contextual revisions. One way of meeting this problem is to apply 'completed context' rather than 'completed sentence' as a criterion, defined as 'semantically meaningful and coherent text', that is, a unit of language which per se is meaningful, which is

not necessarily a sentence. In this sense, a completed sentence⁴ as well as a completed phrase constitute a semantically meaningful and coherent text and can, therefore, form the completed context in which a contextual revision can be made. In the following, this idea is supported by examples from Haugaard (2016) and Haugaard (2018). The following example (A) illustrates a completed sentence to which a premodifier is added.

<p>R0 Entre febrero y marzo, los inversores internacionales retiraron de la deuda española cerca de 25.000 millones de euros, según los datos del Tesoro Público. 'Between February and March, international investors have retired the Spanish debt by approx. 25.000 million Euros, according to State treasury registers.'</p> <p>R1 Entre febrero y marzo, los inversores internacionales retiraron de la deuda española cerca de 25.000 millones de euros, según los <u>últimos</u> datos del Tesoro Público. 'Between February and March, international investors have retired the Spanish debt by approx. 25.000 million Euros, according to <u>the latest</u> State treasury registers.'</p>

Example A

After the text producer has attempted to write the beginning of a sentence four times (at the leading edge of the text), she goes back two sentences (R0) and adds a premodifier, *últimos* (: latest), to that sentence (R1). After the revision, she returns to the leading edge of the text and the beginning of the new sentence. Thus, the addition above is made in a completed sentence and can therefore be defined as a contextual revision according to the LS-taxonomy. The added adjective is a premodifier commonly used in Spanish economic language, which often does not add content to the text that cannot be inferred from the rest of the text. This is also the case here and, accordingly, the revision only applies to form. The revision should be compared to the following revision, made in a sentence that has not yet been ended by a full stop. In this example (B), the text producer interrupts the text production at the leading edge (R0) to add the adjective *net* (: net) in the noun phrase at the leading edge (R1).

<p>R0 Según los últimos números difundidos ayer por el Banco de España en el primer bimestre del año, España sufrió una salida de inversiones de cartera de 'According to the latest figures published yesterday by the Spanish central bank, in the first two months of the year, Spain suffered a fall in portfolio investments [de]'</p> <p>R1 Según los últimos números difundidos ayer por el Banco de España en el primer bimestre del año, España sufrió una salida <u>net</u> de inversiones de cartera de 'According to the latest figures published yesterday by the Spanish central bank, in the first two months of the year, Spain suffered a <u>net</u> fall in portfolio investments [de]'</p>

Example B

Although the noun phrase, *una salida de inversiones de cartera de* (: a fall in portfolio investments [de]), at the time of the revision, is not yet finalized, the part affected by the revision, *una salida de inversiones de cartera*, represents a semantically meaningful text. As it cannot be inferred

⁴ In the following, we rely on a common understanding of 'completed sentence', as a sentence ended by a full stop or another sentence-completing character. >

from the text that it is a net fall, the revision can be interpreted as an addition of content. After this addition, the text producer resumes text production at the leading edge.

The discussion above shows that semantically meaningful contexts can exist independently of a full stop. On this basis, we will argue that (a sentence formally completed by) a full stop or another sentence-completing character cannot be used as a defining feature of a contextual and interpretable revision. Instead, we suggest to define contextual and interpretable revisions as revisions made in a semantically meaningful context which may be, but is not necessarily, a sentence formally completed by a full stop or another sentence-completing character. In sum, it is argued that Lindgren/Sullivan's concepts of 'completed sentence' and 'completed context' should be interpreted as 'semantically meaningful context'.

4.2.2. Contextual revisions and the feature 'preceded and followed by text'

The discussion in section 4.2.1. above concerned the problem regarding one of two defining characteristics of contextual revisions: that they are made in a previously written and completed sentence. As mentioned in section 4.1, the second defining feature of contextual revisions is that they are preceded and followed by text. This feature is also difficult to operationalize, though, because Lindgren/Sullivan (2006b) do not define 'text'.

As proposed in the discussion of 'completed sentence' above, the authors may have refrained from defining 'text', because they rely on the everyday meaning of the word, that is, as a compilation of sentences containing meaningful combinations of words. However, this definition of 'text' is too restrictive to include revisions which are carried out at the beginning of the first sentence of a text and at the end of the last sentence of a text, as these revisions are not preceded or followed, respectively, by 'text' defined in this way. This problem is illustrated by the following examples (C and D).

R0

Mañana, cuando las plazas del Viejo Continente retomen su actividad, el selectivo español arrancará la sesión 1.000 puntos por debajo de lo que lo hizo el pasado 1 de abril.

'Tomorrow, when the European stock exchanges resume their activity, the Spanish stock exchange index will start the session 1,000 points lower than last April 1.'

R1

Mañana, cuando las plazas del Viejo Continente retomen su actividad, el selectivo español arrancará la sesión 1.000 puntos por debajo de lo que despidió marzo.

'Tomorrow, when the European stock exchanges resume their activity, the Spanish stock exchange index will start the session 1,000 points lower than at the end of March.'

Example C

The text producer has completed the third, and last, sentence of the article at the time of the revision (R0). After a pause, she overwrites the last part of the sentence, *lo hizo el pasado 1 de abril* (: lower than last April 1), with *despidió marzo* (: than at the end of March) (R1). This revision is carried out within a completed sentence. In spite of the fact that the revision constitutes the last externalized text at the leading edge and, therefore, is not followed by text, the sentence constitutes a semantically meaningful context, which is a feature of contextual and interpretable revisions (see section 4.2.1 above). Thus, the required analytical context is provided by the completed sentence, which allows an interpretation of the revision as a substitution of form with a different way of expressing the same content. As a result, this revision can be understood as contextual, rather than as pre-contextual. The revision should be compared to the following revision D undertaken in a sentence which was not the last sentence of the article at the time of the revision.

<p>R0</p> <p>Las entidades españolas amasan una cartera de deuda soberana de la Eurozona de 263.300 millones de euros y de ellos, 170.611 son bonos y letras del Estado español, según se desprende de los datos del BCE y el Tesoro. *The Spanish banks purchase a portfolio of Spanish government debt from the Euro zone of 263.300 million Euros and, out of these, 170.611 are short term government bonds and Spanish Treasury bills, according to figures from the European Central Bank and the Ministry of Economic Affairs.*</p> <p>R1</p> <p>Las entidades españolas amasan una cartera de deuda soberana de la Eurozona de 263.300 millones de euros y de ellos, 170.611 son bonos y letras del Estado español, según se desprende de los datos del BCE y <u>el emisor español</u>. *The Spanish banks purchase a portfolio of Spanish government debt from the Euro zone of 263.300 million Euros and, out of these, 170.611 are short term government bonds and Spanish Treasury bills, according to figures from the European Central Bank and <u>the Spanish Central Bank</u>.*</p>

Example D

The text producer has completed a sentence, which in itself is a contextual revision as it is added between two sentences (R0). After a pause, the last word of this sentence, *Tesoro* (: the Ministry of Economy) is overwritten with *emisor español* (: the Spanish central bank). Because the sentence in which the revision is made in itself is a contextual addition and, therefore, followed (and preceded) by several completed sentences, it is important to stress that it is not the fact that the revision is followed by text – the already completed sentences, which would qualify it as a contextual revision – but the semantically meaningful context in which the revision occurs. Thus, examples C and D show that text following a revision should not be a defining feature in itself. The examples provided by Lindgren/Sullivan (2006b: 174-175) indicate that the authors treat similar revisions as contextual. However, as the examples are described in isolation and not in context, it is not apparent whether the revisions are similarly made at the end of the last sentence of the overall externalized text.

It follows from the discussions of this section that also the defining feature of a contextual revision as being preceded by text is problematic and influences the way revisions are classified at the beginning of a semantically meaningful context. If a revision is made in the first sentence of the text, it is not preceded by text and the revision cannot be classified as contextual, according to Lindgren/Sullivan's criteria. On the contrary, if the sentence is located elsewhere in the text, the revision is preceded by text allowing for an interpretation as contextual. Hence, identical revisions will be categorised differently depending on whether the sentence in which the revision is undertaken appears as the first sentence in a text-production process or in a subsequent sentence. Consequently, the fact that a revision is preceded by text cannot be a defining feature in itself.

In conclusion, we argue that the feature of being preceded or followed by text should not be used as a defining criterion of contextual revisions. The exclusion of this criterion allows for an interpretation of contextual revisions as revisions made in a semantically meaningful context located anywhere in the text, including at the beginning of the first sentence of a text and at the end of the last sentence of a text, i.e. the leading edge.

4.2.3. Additions at the end of a completed sentence – contextual or pre-contextual revisions?

The two examples above (C and D) both illustrate a replacement of one element by another within the context of a completed sentence. When the text producer chooses to add an element at the end of an otherwise completed sentence, the interpretation and classification of the revision are more complicated if the revision involves the deletion of the full stop followed by the transcription of text before the full stop is reinserted. The problem is illustrated by example E.

R0

Curiosamente, esta cifra es similar a la desinversión realizada por los inversores extranjeros en deuda Española en el mes de marzo, lo que revela que el débil sistema financiero español está siendo el principal Salvador del Tesoro en las subastas de deuda.

‘Curiously, this figure corresponds to foreign investors’ divestment of Spanish debts in the month of march, which reveals that the weak Spanish financial system is the major saviour of the Ministry of Economic Affairs in the debt auctions.’

R1

Curiosamente, esta cifra es similar a la desinversión realizada por los inversores extranjeros en deuda Española en el mes de marzo, lo que revela que el débil sistema financiero español está siendo el principal Salvador del Tesoro en las subastas de deuda a cambio de obtener unas jugosas rentabilidades que le permitirán mejorar su cuenta de resultados, muy dañada por la caída del negocio en España y por el esfuerzo que los bancos están realizando para cumplir con los requisitos de la nueva reforma financiera.

‘Curiously, this figure corresponds to foreign investors’ divestment of Spanish debts in the month of march, which reveals that the weak Spanish financial system is the major saviour of the Ministry of Economic Affairs in the debt auctions, in return for juicy returns, which can help improving their bottom line, very much affected by the deceleration of business in Spain and by the effort that the banks make to meet the demands of the new financial system.’

Example E

The text producer has ended a sentence by a full stop, but after a short pause, in one coherent sequence, she deletes the full stop, adds a long prepositional phrase and ends the sentence again by a full stop. This revision is carried out *both* in a semantically meaningful context, which makes it contextual, *and* at the leading edge of the text, which makes it pre-contextual, according to Lindgren/Sullivan’s (2006b) criteria. In this sense, the revision is situated at the interface between contextual and pre-contextual revisions. However, on the basis of the conclusions of sections 4.2.1-4.2.3, we will argue that this revision is made in a semantically meaningful context and should be classified as a contextual revision in order to obtain a stringent analysis.

4.2.4. Summing up

Based on the above discussions, it appears that the defining features that Lindgren/Sullivan (2006b) set for contextual revisions reflect an unclear and possibly restrictive understanding of contextual revisions, as revisions made in the middle of a completed sentence (see, for example, example A above where *últimos* is added to a sentence completed by a full stop). As illustrated above, this understanding not only eliminates revisions which are comparable to those which can be identified by the LS-taxonomy, but also excludes interpretable and interesting revisions that can enhance our understanding of revisions made during writing. Therefore, we suggest to define Lindgren/Sullivan’s concepts ‘completed sentence’ and ‘completed context’ as ‘semantically meaningful context’. Accordingly, contextual and interpretable revisions are defined as revisions which are made in a semantically meaningful context. A semantically meaningful context may be, but is not necessarily, a sentence formally completed by a full stop or another sentence-completing character. Moreover, we argue that text preceding and following a revision should not be a defining feature in itself, as this will entail a heterogeneous analysis. This implies that the semantically meaningful context in which contextual and interpretable revisions are made can be located anywhere in the text, also at the leading edge. Lastly, it is a further advantage of applying semantically meaningful context as a criterion that it can help disambiguating the classification of additions in completed sentences at the leading edge.

4.2.5. Pre-contextual revisions

As pointed out in the outline of external revisions of section 4.1, Lindgren/Sullivan (2006b: 159) define pre-contextual revisions as changes in the text just written at the leading edge with context only at the left of the revision. Hence, pre-contextual revisions are made when the text producer, in his/her work with content and form of the emerging text becomes aware that what has just been transcribed or is currently being transcribed requires adjustment. The lack of a semantically meaningful context, which often characterises pre-contextual revisions, implies that the effect of the revisions on the text is difficult – or even impossible – to interpret on the basis of keystroke logging. This is illustrated by the following example.

<p>(R0) La Bolsa española cerró ayer el peor abril en 25 años y su peor mes en *The Spanish Stock Exchange closed yesterday the worst April in 25 years and its worst month in*</p> <p>(R1) La Bolsa española cerró ayer el peor abril en 25 años y su peor mes <u>desde noviembre de 2010</u> *The Spanish Stock Exchange closed yesterday the worst April in 25 years and its worst month <u>since November 2010</u>*</p>

Example F

The text producer is working on the first sentence of her article (R0). As soon as she has transcribed the preposition *en* (: in) at the leading edge, she replaces it by another preposition, *desde* (: since), while also transcribing the complement of the prepositional phrase, *noviembre de 2010* (: November 2010) (R1).

In this example, the prepositions, *en* and *desde*, depend on the complement not yet transcribed at the time of revision. There is, thus, no semantically meaningful context at the time of the revision on the basis of which it may be determined whether the revision reflects a change in the content which is to be transcribed, or whether it affects only the form of the text. Hence, the interpretation of the effect of this revision, like the majority of revisions made in semantically non-meaningful contexts, would require triangulation of other methods, such as a verbal protocol from the writer, e.g. think aloud protocols or stimulated recall.

The following discussion zooms in on revisions that lie at the interface between contextual and pre-contextual revisions, which we will call pseudo pre-contextual revisions. These revisions are divided into two subtypes: omission and replacement at the leading edge, on the one hand, and retranscription at the leading edge, on the other.

4.2.5.1. Pseudo pre-contextual revisions – omissions and replacements at the leading edge

From the discussion of sections 4.2.1 and 4.2.2 above, it appears that contextual and interpretable revisions are considered to be revisions made in a semantically meaningful text that may be, but is not necessarily, a completed sentence. On the basis of this definition, also a phrase can provide a semantically meaningful context in which a contextual and interpretable revision can occur. In rare cases, a semantically meaningful text appears at the leading edge of the text where new text is being transcribed and, accordingly, shares common features with pre-contextual revisions shaping the emerging text. However, because the text which is subject to revision constitutes a semantically meaningful text, the revision can be interpreted. This applies to revisions in which a part of a semantically meaningful phrase is left out and revisions which involve the replacement of one word or phrase with another. These two categories are illustrated by examples G and H below.

R0

Esto es, 3.127 millones de euro

'This is 3,127 billion Euros'

R1

Esto es, 3.127 millones.

'This is 3,127 billion.'

Example G

In example G, the text producer pauses before the noun phrase *millones de euro(s)* (: billion euro(s)) has been fully transcribed (R0) and deletes the unfinished complement of the phrase *de euro* (: euro) at the leading edge and sets a full stop (R1). In this sense, the complement of the phrase (*de euro(s)*, which restricts the meaning of the referent of the head, (*millones* (: billion)), is left out. Yet, this information is already apparent from the text written thus far (not part of this example, though), and it is, therefore, possible to interpret the effect of the revision as affecting only the form of the text.

The example above illustrates an interpretable omission in a semantically meaningful context at the leading edge. The next example H, illustrates a replacement.

R0

Ambas entidades coincidieron también en criticar la manera en la que el Banco de España y el Minis

'Both banks also criticized the way in which the Spanish Central Bank and the Minis'

R1

Ambas entidades coincidieron también en criticar la manera en la que el Banco de España y Economía están'Both banks also criticized the way in which the Spanish Central Bank and the (Ministry of) Economic Affairs [están]'

Example H

The text producer pauses in the middle of the transcription of the noun, *Mini(sterio)* (: Mini(stry)). She, then, deletes *Minis* replacing it by *Economía* (: (The ministry of) economic (affairs)) while also transcribing the verb, *están*. In this example, the text already written contributes to an understanding of the revision. In a previous sentence, *Ministerio de Economía* (: Ministry of Economic Affairs) has been mentioned and therefore the revision can be interpreted as a replacement of *Minis(terio de Economía)* with the shorter version, *Economía*, that is, a lexical variation commonly used with a reference to the *Ministerio de Economía*.

It is a common characteristic of the interpretation of the two examples above that it is informed by the text written thus far although the contexts in which they are made are semantically non-meaningful.

4.2.5.2. Pseudo pre-contextual revisions – retranscription at the leading edge

There is yet another interpretable revision type at the leading edge of the text. This type involves deletion and retranscription of (part of) the deleted text. In other words, deleting text is used as a means to reach the point in the text where the actual revision is to be carried out. A systematization of these seemingly pre-contextual revisions suggests that deleting and retyping at the leading edge of the text are used for omitting and replacing previously (partially) written text as well as for adding new text, which is exemplified by the following three examples (I, J and K respectively).

<p>R0 En los últimos 30 días, el Ibex 35 ha perdido 1.000 puntos y arra 'In the last 30 days, Ibex 35 has decreased by 1.000 points and [arra]'</p> <p>R1a En los últimos 30 días, el Ibex 35 ha perdido 1.000 puntos y 'In the last 30 days, Ibex 35 has decreased by 1.000 points and' []</p> <p>R1b En los últimos 30 días, el Ibex 35 ha perdido 1.000 puntos y <u>mañana arrancará el mes de may</u> 'In the last 30 days, Ibex 35 has decreased by 1.000 points and <u>tomorrow the month of May starts</u>'</p>

Example I

In example I, the text producer is working on the second sentence of her article (R0). At the leading edge of the text, the beginning of the verbal, *arra(ncará)* (: starts) is deleted before it is fully transcribed (R1a). Then, the adverb of time, *mañana* (: tomorrow), is transcribed, after which the entire verbal, *arra(ncará)* is retyped along with the next unit of language (R1b). The revision and the subsequent text production are carried out in one consecutive sequence and the retranscription of the verbal *arra(ncará)* suggests an interpretation of the revision as a content addition of *mañana*. Example J illustrates omission of previously written text.

<p>R0 En los últimos 30 días, el Ibex 35 ha perdido 1.000 puntos y mañana arrancará el mes de may 'In the last 30 days, Ibex 35 has decreased by 1.000 points and tomorrow the month of May starts'</p> <p>R1a En los últimos 30 días, el Ibex 35 ha perdido 1.000 puntos y mañana arrancará [] 'In the last 30 days, Ibex 35 has decreased by 1.000 points and tomorrow starts'</p> <p>R1b En los últimos 30 días, el Ibex 35 ha perdido 1.000 puntos y mañana arrancará [] <u>mayo</u> (hoy es un día festivo para las Bolsas europeas) 'In the last 30 days, Ibex 35 has decreased by 1.000 points and tomorrow [] <u>May</u> starts (today is an official holiday in the European stock exchanges)'</p>

Example J

In this example, the text producer works on the same sentence as in example I. She stops transcribing before the word *may(o)* (: May) is fully transcribed (R0), hesitates and then deletes the last part of the sentence *el mes de may* (: the month of May) (R1a) whereupon she retypes *may(o)*, lingers and continues the text production. The retranscription of *may(o)* licenses the interpretation of the redundant *el mes de* in terms of omission. This revision, therefore, affects only the form of the text.

Examples I and J above illustrate an addition and an omission, respectively, by means of deleting and retyping. The following example illustrates a replacement.

<p>R0 Según los últimos datos de la institución monetaria, la banca española incrementó en el mes de marzo su exposición a bonos <i>sobe</i> ‘According to the latest figures from the European Central Bank, in May, the Spanish banks have increased their exposure to [sobe] bonds’</p> <p>R1a Según los últimos datos de la institución monetaria, la banca española incrementó en el mes de marzo su exposición a [] ‘According to the latest figures from the European Central Bank, in May, the Spanish banks have increased their exposure to []’</p> <p>R1b Según los últimos datos de la institución monetaria, la banca española incrementó en el mes de marzo su exposición a <u>deuda soberana</u> en más de 20.000 millones de euros. ‘According to the latest figures from the European Central Bank, in May, the Spanish banks have increased their exposure to <u>sovereign debts</u> with more than 20.000 million Euros’</p>

Example K

The text producer ceases transcribing in the middle of an adjective, *sobe(ranos)*. After a short hesitation, she replaces *bonos sobe* (: bonds [sobe]) by *deuda sobe(rana)* (: sovereign debts) and adds a prepositional phrase, which is ended by a full stop. In this example, the retranscription of the adjective *sobe(rana)* indicates that the change from *bonos* to *deuda* is a replacement of a term by its hypernym, thus affecting the content.

4.2.2. Summing up

The examples of sections 4.2.5, 4.2.5.1. and 4.2.5.2. show, on the one hand, that, due to the fact that the context of pre-contextual revisions is semantically incomplete, the effect on the text of the revisions cannot be interpreted. On the other hand, pseudo pre-contextual revisions, that is, revisions in which a part of a semantically meaningful phrase is left out (G), or revisions which involve the replacement of a word/phrase with another (H), or the retranscription of (part of) a deleted text (I, J and K), may lend themselves to interpretation.

4.3. A reassessment of the LS-taxonomy

While accepting the basic distinctions on which the LS-taxonomy is founded, it is argued that the discussion above calls for adjustments, both of the conceptualization of the taxonomy itself and of the definition of some of the basic concepts. The idea of a taxonomy presupposes that phenomena or concepts are discrete entities that can be grouped into closed classes and that there exists a set of clear criteria according to which classification can be carried out. The study reported in this article shows that external revisions do not, in all cases, allow for a straightforward classification.

We suggest to replace the taxonomy of external revisions with a continuum of semantically meaningful context, with semantically meaningful context and semantically non-meaningful context as opposite poles. Different types of context can be located in the intermediate zone between the two poles, depending on their completeness.

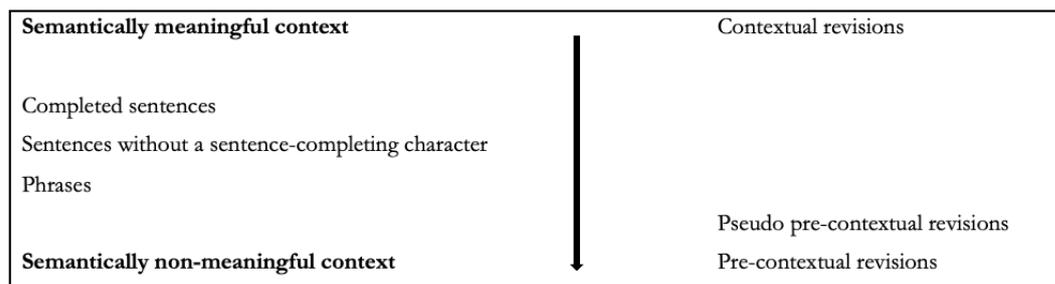


Figure 3: Continuum of semantically (non-)meaningful context

In this continuum, contextual revisions made in a completed sentence (examples A, C, D and E) are placed at one end⁵, followed by revisions in semantically meaningful contexts which are potentially less complete than a sentence ended by a sentence-completing character, such as semantically meaningful sentences without a sentence-completing character and semantically meaningful phrases (example B, G and H). It is important to note that examples G, H, I, J and K were classified as pre-contextual revisions in the analysis of section 4.2.5, but the idea of a continuum takes into account the fact that they have features in common with contextual revisions, lending themselves to interpretation. At the other end of the continuum (example F), we find pre-contextual revisions which are characterized by the semantically non-meaningful contexts in which they are made (example F). Only the effect of revisions made in semantically meaningful contexts can be interpreted solely on the basis of keystroke logging.

5. Concluding remarks

When Lindgren and Sullivan developed the LS-taxonomy, it filled out a gap in the investigation of revision by providing a framework with a process-oriented and online approach to the study of external revisions of form and content. However, when exposed to new data, the LS-taxonomy exhibits weaknesses when it comes to classifying external revisions on the basis of the dichotomous distinction between contextual and pre-contextual revisions. In order to improve the explanatory power of the model, we advocate a dynamic approach in which revisions are classified on a continuum of semantic completeness of the context. In other words, it is suggested to replace the taxonomy of external revisions by a continuum of semantically meaningful context at one end with semantically meaningful context and semantically non-meaningful context at the other end. Different types of context exhibiting varying degree of completeness can be located in the intermediate zone between the two poles. With the exception of a particular group of pseudo pre-contextual revisions, only the effect of revisions made in semantically meaningful contexts can be interpreted solely on the basis of keystroke logging. By introducing an interpretation as to whether the revision is carried out in a semantically meaningful context, the analytical framework suggested (in this article) distances itself from the more categorical classification of the location of revisions at the leading edge or in the text already written of Lindgren/Sullivan. This allows for a stringent and homogenous analysis of contextual and interpretable revisions located anywhere in the text and facilitates a systematisation of the contexts in which the effect of pseudo pre-contextual revisions can be interpreted based on keystroke logging.

⁵ The entire completed text in which sentences are embedded reinforces the completeness of the individual sentences in which a revision can be made. However, as the highest discursive level at which a revision can be made is the individual sentence, we chose this concept for one of the extremes of the continuum.

6. References

- Alamargot, Denis/Chanquoy, Lucile 2001: *Through the Models of Writing*. Dordrecht: Kluwer Academic Publishers.
- Baaijen, Veerle M./ Galbraith, David/Glopper, Kees de 2012: Keystroke Analysis: Reflections on Procedures and Measures. In *Written Communication* 29(3), 246–277.
- Bridwell, Lillian S. 1980: Revising strategies in twelfth grade students' transactional writing. In *Research in the Teaching of English* 14(3), 197-222.
- Chanquoy, Lucile 2009: Revision processes. In Beard, Roger/Myhill, Debra/Riley, Jeni/Nystrand, Martin (eds.), *The SAGE Handbook of writing development*. London: SAGE Publications Ltd, 80-97.
- Chenoweth, N. Ann/Hayes, John R. 2001: Fluency in Writing: Generating Text in L1 and L2. In *Written Communication* 18, 80-98.
- Chenoweth, N. Ann/Hayes, John R. 2003: The Inner Voice in Writing. In *Written Communication* 20, 99-118.
- Faigley, Lester/Witte, Stephen 1981: Analyzing Revision. In *College Composition and Communication* 32(4), 400-414.
- Fitzgerald, Jill 1987: Research on revision in writing. In *Review of Educational Research* 57(4), 481-506.
- Flower, Linda/Hayes, John R. 1981: A cognitive process theory of writing. In *College Composition and Communication* 32(4), 365-387.
- Freedman, Sarah W. 1985: Introduction. In Freedman, Sarah W. (ed.), *The Acquisition of Written Language: Response And Revision*. Nordwood, New Jersey: Ablex Publishing Corporation, x-xv.
- Haugaard, Rikke Hartmann 2016: *Hvordan skriver og reviderer journalister? Et eksplorativt og kvalitativt casestudie af journalistisk tekstproduktion med fokus på revision af indhold og sprogligt udtryk ved det spanske dagblad El Mundo*. PhD thesis, Aarhus university.
- Haugaard, Rikke Hartmann 2018: Journalistic News Writing: A Case Study on Revisions of Content and Form. In *Fachsprache. Journal of Professional and Scientific Communication* 40(3-4), 122-139.
- Hildick, Wallace 1965: *Word for word: The rewriting of fiction*. London: Faber and Faber.
- Hill, Charles A/Wallace, David L/Haas, Christina 1991: Revising On-Line: Computer Technologies and the Revising Process. In *Computers and Composition – An International Journal* 9(1), 83-109 [online]. http://computersandcomposition.candcblog.org/archives/v9/9_1_html/9_1_6_Hill.html (accessed 12 April 2018).
- Humes, Ann 1983: Research on the Composing Process. In *Review of Educational Research* 53(2), 201-216.
- Kaufert, David. S/Hayes, John R/Flower, Linda 1986: Composing Written Sentences. In *Research in the Teaching of English* 20(2), 121-140.
- Kollberg, Py/Severinson Eklundh, Kerstin 2001: Studying writers' revising patterns with s-notation analysis. In Olive, Thierry/Levy, C. Michael (eds.), *Contemporary tools and techniques for studying writing*. Dordrecht/Boston/London: Kluwer academic publishers, 89-104.
- Krogh, Ellen 2003: *Et fag i moderniteten. Danskfagets didaktiske diskurser*. PhD dissertation, Dansk Institut for Gymnasiepædagogik, Syddansk Universitet.
- Lindgren, Eva 2005: *Writing and Revising. Didactic and Methodological Implications of Keystroke Logging*. PhD thesis, Skrifter från moderna språk 18, Department of Language Studies, Umeå University.
- Lindgren, Eva/Sullivan, Kirk P. H. 2006a: Writing and the analysis of revision: An overview. In Sullivan, Kirk P. H./Lindgren, Eva (eds.), *Computer Keystroke Logging and Writing*. Oxford: Elsevier, 31-44.
- Lindgren, Eva/Sullivan, Kirk P. H. 2006b: Analyzing on-line revision. In Sullivan, Kirk P. H./Lindgren, Eva (eds.), *Computer Keystroke Logging and Writing*. Oxford: Elsevier, 157-188.
- Matsuhashi, Ann 1987: Revising the text and altering the plan. In Matsuhashi, Ann (ed.), *Writing in real time. Modeling text production processes*. Norwood, NJ: Ablex Publishing Corporation: 197-223.
- Murray, Don 1978: Internal revision: A process of discovery. In Cooper, Charles R./Odell, Lee (eds.), *Research on composing. Points of departure*. National Council of Teachers of English. Illinois, 85-104.
- National Assessment of Educational Progress (NAEP) 1977: *Write/rewrite: An assessment of re-vision skills: Selected results from the second national assessment of writing*. ERIC Document Reproduction Service No. ED 141 826.
- Nold, Ellen 1979: Alternatives to mad-hatterism. In McQuade, Donald. (ed.), *Linguistics, stylistics, and the teaching of composition*. Akron, OH: University of Akron, 103-117.
- Nystrand, Martin 2001: Cultural support for empirical research on writing. In Coppock, Patrick (ed.), *The semiotics of writing: Transdisciplinary perspectives on the technology of writing*. Turnhout: Brepols, 115-136.
- Perl, Sondra 1979: The composing process of unskilled college writers. In *Research in the Teaching of English* 13(4), 317-336.

- Perrin, Daniel 2003: Progression analysis (PA): investigating writing strategies at the workplace. In *Journal of Pragmatics* 35, 907–921.
- Perrin, Daniel 2013: Investigating language and the media: The case of newswriting. In *AILA Review* 26, 57–78.
- Rohman, D. Gordon/Wlecke, Albert O. 1964: *Pre-Writing: The Construction and Application of Models for Concept Formation in Writing*. U.S. Department of Health, Education, and Welfare Cooperative Research Project No. 2174. East Lansing: Michigan State.
- Severinson Eklundh, Kerstin/Kollberg, Py 1996: A computer tool framework for analyzing online revisions. In Levy, C. Michael/Ransdell, Sarah (eds.), *The Science of Writing: Theories, Methods, Individual Differences, and Applications*. Mahwah, NJ: Lawrence Erlbaum, 163-188.
- Sommers, Nancy 1980: Revision strategies of student writers and experienced adult writers. In *College Composition and Communication* 31(4), 378-388.
- Stallard, Charles K. 1974: An analysis of the behavior of good student writers. In *Research in the Teaching of English* 8, 206-218.
- Stevenson, Marie/Schoonen, Rob/Glopper, Kees de 2006: Revising in two languages: A multi-dimensional comparison of online writing revisions in L1 and FL. In *Journal of Second Language Writing* 15, 201–233.
- Van Hout, Tom 2010: *Writing from Sources: Ethnographic Insights into Business News Production*. PhD thesis. Department of Language and Communication, Ghent University.
- Van Waes, Luuk/Schellens, Peter Jan 2003: Writing profiles: the effect of the writing mode on pausing and revision patterns of experienced writers. In *Journal of Pragmatics* (35), 829-853.
- Witte, Stephen P. 1985: Revising, Composing Theory, and Research Design. In Freedman, S. W. (ed.), *The Acquisition of Written Language: Response and Revision*. Norwood, New Jersey: Ablex Publishing Corporation, 250-284.
- Witte, Stephen P. 1987: Pre-Text and Composing. In *College Composition and Communication* 38(4), 397-425.