Journalistic News Writing: A Case Study on Revisions of Content and Form

Rikke Hartmann Haugaard

Abstract News products provide a major part of the foundation of what we know about the world in which we live. However, we lack empirical knowledge about the process of writing news texts, e. g. knowledge about the choices made by journalists as to *what* to communicate and *how* to communicate it. This paper aims to contribute to filling this research gap by reporting on a case study of some of the decisions journalists make as regards content and linguistic form when composing news articles. More specifically, the study investigated the revision practices of three journalists during text production at a Spanish newspaper, as these revisions yield insights into the progression of the text and thus contribute to our understanding of how journalists work with the content and linguistic form of a text. Results indicate that journalists' revisions are related to form markedly more often than to content (approx. three to four times more often). Moreover, revision type (e. g. addition, omission and substitution) and effect on the text (content or form) seem to suggest two writing phases serving different overall purposes; producing (more) text for the newspaper article in the first phase and evaluating, and especially reducing the length of the article in the second phase.

Keywords journalistic writing and revising, democratisation of knowledge, professional text production, workplace writing, online revisions, revisions of content and form, keystroke logging, qualitative research

1 Introduction

According to the American Press Institute, the "purpose of journalism is [...] to provide citizens with the information they need to make the best possible decisions about their lives, their communities, their societies, and their governments" (American Press Institute 2017). Hence, a journalist's basic task is to keep citizens informed about what is currently happening in society and in the world at large, thus empowering the citizens to make the best possible choices. This democratising role of journalism implies that news media shape the settings in which we discuss events and get information about ourselves and the world we live in (Hartley 2011: 21). Consequently, news journalism possesses an orchestrating, manipulating power regarding the public debate (Willig 2011: 16; cf. also Jakobs/Perrin 2014a: 1). For this reason, and because journalistic products provide a large part of the foundation for what we know about the world we inhabit, it becomes important to examine how news articles are written.

The last stage before a news article reaches the consumers is the very process of writing, in which the journalist, among other things, makes deliberate choices concerning the content and form of the news, i. e. choices pertaining to *what* to communicate, on the one hand, and

Zitiervorschlag / Citation:
Haugaard, Rikke Hartmann (2018): "Journalistic News Writing: A Case Study on Revisions of Content and Form." *Fachsprache. Journal of Professional and Scientific Communication* 40.3–4: 122–140.

how to communicate it, on the other. In other words, these are choices made to ensure that the story is understood by the consumers, thus empowering them and consequently creating social value[1]. Typically, studies in news production do not look into this journalistic writing process. However, a few studies have dealt specifically with journalistic text production, e. g. Pitts (1982), Schumacher/Scott/Klare et al. (1989), Perrin (e. g. 2001) and Van Hout (2010), but our empirical knowledge about journalists' choices of *content* and *form* in the production of news articles is still very limited. The current paper aims to contribute to filling this research gap by investigating the content-form dichotomy in journalistic text production.

In terms of theory, this study is rooted in the process-oriented, cognitive writing research concerned with investigating and understanding how texts are written. As a cognitive activity, writing is perceived as a recursive activity drawing on hierarchically organised cognitive processes which are, somewhat simplistically, the cognitive elements that plan and generate, formulate, evaluate, and revise content and form if necessary.[2] Writers use these processes actively and in different combinations during text production "to progressively create a text that meets their conception of topic, task and audience" (Lindgren/Sullivan 2006a: 32). The interaction between planning, generating, formulating, evaluating and revising results in a recursive process as text producers elaborate on their texts, giving up existing ideas and structures in favour of new ones (Lindgren/Sullivan 2006a: 32). The recursivity and the hierarchical structure imply that choices regarding the content and form of a text are continuously being evaluated and revised while the text is being produced. Accordingly, revision constitutes an inherent part of text production. In the study, the "revision strand" was extracted for separate examination, precisely because revision provides insights into the progression of a text and thus contributes with a significant element to understanding how journalists shape the content and form of the evolving text.

Thus, focusing on the content-form dichotomy in journalistic writing and revising, the present paper reports on an observational study (Haugaard 2016), aiming at gaining insight into some of the decisions concerning content and form which journalists made while producing a text. The study explored different aspects of revision occurring during three professional text producers' ordinary writing practices as they unfolded in everyday life in an editorial office of a major Spanish newspaper.[3]

---

1 The why question, which is the reasoning behind the journalists's choices, is also highly relevant. Nevertheless, an investigation of this aspect is beyond the scope of this article.

2 From a more context-oriented perspective, these cognitive processes are embedded in and influenced by a number of contextual factors, which affect both the writing process and the product. This line of research is demonstrated in the work of various scholars, e. g. Van Hout's (2010) ethnographic approach to the study of intertextuality in newswriting in the workplace, Perrin's (2013) "Linguistics of Newswriting" and Leijten/Van Waes/Schriver et al.'s (2014) comprehensive model of skilled professional text production.

3 Research on revision has been engaged both in classifying the textual changes that writers make to a text and modelling the complex cognitive and/or social processes that are assumed to precede or influence these changes. Thus, the term *revision* has been used with reference to process and textual changes alike (e. g. Fitzgerald 1987: 483, Fredmann 1985: xi, Alamargot/Chanquoy 2001: 100) (see Haugaard 2016: 52 for a brief summary of other terms used). The object of study of this paper is the textual changes made in the text during writing.

For each journalist, the study investigated the characteristics of the revisions of content and form separately. This means that the study examined *time of occurrence* during the writing process, *revision type,* such as addition, omission and substitution, and the possible *relation* between timing and revision type. Moreover, the study analysed the *distribution* of revisions on content and form and the *differences* between and *similarities* shared by the three journalists.

In the following, the research design will be described, the analytical framework of the study will be introduced, and the results will be presented and discussed.

2 Research design

Research on writing is an eclectic field comprising a great variety of disciplines, traditions, approaches to and perspectives on text production (Jakobs/Perrin 2014b: 27). Even so, many researchers consider the composing process to be both a cognitive and a situated activity (e. g. Schultz 2006: 368). Recognising that text production does not take place in a vacuum, and that the writing process and, hence, the genesis of the text are influenced by a number of external factors affecting both process and product, the study combined a cognitive and a contextual approach (cf. also Perrin e. g. 2001, 2003, 2013, Perrin/Ehrensberger-Dow 2006, Van Hout 2010, Leijten/Van Waes/Schriver et al. 2014) in the examination of naturally occurring composing processes embedded in everyday life at a workplace.

Using case studies is highlighted as a suitable method for investigating the peculiarities and complexities of a phenomenon (Stake 1995: xi) within its real-life context (Yin 2003: 13). Furthermore, multiple cases often accentuate complementary aspects of a phenomenon and can therefore strengthen the accuracy and validity of the analysis (Neergaard 2007: 22). Based on the research interest of the study, i. e. identifying and comparing specific characteristics of journalists' revisions, a qualitative and exploratory multiple case study was chosen to explore and thoroughly describe authentic journalistic writing and revising in their natural context. More specifically, the study investigated three instances of text production processes embedded in the journalistic everyday life at the business section of the Spanish national newspaper *El Mundo*. The participants were selected using strategic and convenience sampling, ensuring a sample of journalists with suitable qualities and a willingness to participate in the study[4] (Thagaard 2012: 55 f.). As regards level of experience, the participants had been working at the business section for between two and nine years (cf. Table 1); two of the participants had at least a couple of years of experience within their subject area, while one participant (P1) had recently changed her subject area due to a reorganisation of the section and had only a few months of experience within this new field. This heterogeneity in relation to years of employment at the section and expertise within subject areas was not thought to be problematic, as all three participants had either experience from other sections and subject areas within *El Mundo* (P1, P2) and/or experiences prior to their employment at *El Mundo*, including other media houses with similar (P3) and other (P1, P3) subject areas. Moreover, they all researched their own articles and wrote the final version themselves. Accordingly, they all were "experienced" journalists "with length, depth, and/or breadth of experience" (Perrin/Ehrensberger-Dow 2006: 320). Before the study, the participants had been informed as to how the study would proceed.

[4] Four journalists agreed to participate in the study, but due to technical problems during observations, the study only included empirical material from three articles.

Adhering to a mixed methods approach placing the research agenda at the centre (Johnson/Onwuegbuzie/Turner 2007) and in order to present a description as comprehensive as possible of the revisions made during the three composing processes, the study applied a combination of qualitative and quantitative methods (Creswell 2014: 4, Dam-Jensen/Heine 2009: 11 f.), i. e. the multi-method approach of Progression Analysis (PA) which aims at collecting and analysing writing in natural workplace settings at three different levels: the macro level, i. e. the situational context of the writing process by means of interviews and observations, the meso level, i. e. the evolving text during writing by means of keystroke logging[5, 6], and the micro level, i. e. the writers' consciously applied strategies by means of cue-based retrospective verbalisations (e. g. Perrin 2003). More specifically, the present study made use of keystroke logging and participant observation during the writing process and retrospective interviews immediately following the process giving insights into the journalists' revisions of content and form. In the present study, the macro level and the micro level feed into the analysis and interpretation of the meso level, that is: revisions of content and form.

During the elaboration of the research design, pilot tests were conducted primarily to qualify the participant observation and the retrospective interview. The observation was based on a template with predefined time slots (0–59) that allowed both specified observations (e. g. keyboard activity [insertion or deletion], mouse movements, the use of notes, etc.) and comments to be recorded (e. g. talks to colleague, visits to the toilet, etc.). In this sense, the participant observations served, to a great extent, as a triangulation of the log files. Anticipating that the time for the retrospective interview would be very limited, the interview guide concentrated primarily on the overall process and the how and why of the genesis of the text, and focused on specific activities or challenges that the journalists might identify as relevant to speak about.

Data generation

Striving for homogeneous composing processes in terms of non-participant-dependent variables such as article length and time of writing, the aim was to study the production of the prioritised news article of the day, which has a length of approximately 5000 characters, including spaces. However, for various reasons, this was not possible in all cases, and consequently the texts vary in length (cf. Table 1). All articles were written during the hours immediately before deadline and were printed the following day in the business section of the newspaper, thus addressing the same potential readers.

---

5 A keystroke logger is a type of software that records the writing activity as writers compose on a computer, i. e. which keys are activated, what is deleted and how the cursor moves while the text is being written. The software records the exact time of each activity, including the pauses between them, giving access to all the various stages a text passes through before reaching its final state. The recording is saved in a log file which can be used to analyse various aspects of the text production process.

6 At this level, other computer loggings such as screen recordings and eyetracking have also been used (e. g. Ehrensberger-Dow/Perrin 2013: 77).

Table 1: Participants' subject areas, years of employment at the business section and article details

	Participant 1 (P1)	**Participant 2 (P2)**	**Participant 3 (P3)**
Subject area	banking[7]	labour market	financial market
Years of employment	9	2	4
Date and time of text production	26 April 2012 approx. 18.43–20.45[8] hrs	27 April 2012 approx. 19.55–21.07 hrs	30 April 2012 approx. 19.35–20.55 hrs
Number of words	533	653	746

Prior to the study and after being security-cleared, the keystroke logging software, Translog (Carl 2012), was installed on the participants' personal computers. When a participant was about to write an article, she was introduced to the software, and the recording was then started simultaneously with the participant observation.[9]

To ensure fruitful retrospection, the retrospective interviews were conducted immediately after the text production process and were only interrupted by the normal obligations that apply when an article is submitted for print. The retrospective interviews were conducted as semi-structured interviews in front of the participants' computers. With very few exceptions, the retrospective interviews did not provide insights that could inform the analysis of the revisions. However, the retrospection proved useful for creating insight into the normal workflow of journalists and into the sources and tools they used during the composing processes.

In addition to the completed article, the dataset comprised three process protocols, each containing a keystroke log, an observation protocol, a retrospective interview, source material including handwritten notes, a search history and the tools used during text production. As keystroke logging provides a detailed and nuanced picture of how the articles come into being, as well as nuanced insight into the revisions of content and form, the keystroke logs were the primary data source, and the analysis of these was informed by the remaining empirical material, particularly the observation protocols and the retrospective interviews. In preparation for the analyses, the log files were synchronised with the observation protocols and the retrospective interviews and merged into one document (cf. Table 2).

7 Changed from transport, turism and infrastructure.

8 P1 had a downtime period of about 45 minutes approx. halfway through the process during which she didn't produce text. Moreover, during this period and about 50 minutes into the process, the editors changed the spot for the article in the newspaper moving it from the front page to a location inside the section, which resulted in less space.

9 Afterwards, the Translog recordings were converted into another file type to be analysed by another type of keystroke logging software, Inputlog (Leijten/Van Waes 2013), with different and more sophisticated qualities (see Haugaard 2016 for the reasoning and a thorough discussion of the two softwares).

Table 2: Synchronisation and merging of empirical material

Time – log file	**Time – observation**	**Logging activity**	**Description**	**Revision**	**Paragraph no. / sentence no.**	**Written text**
Pause: 15085 ms	07,16		Observation: Uses little red book Retrospective interview: Checks up on date		1/3	
7,29,94-7,46,47		[▼][▲]despidió marzo.[▼][▲][▼][▲][▼][▲][_][o] _nivel en el [▼][▲]	The cursor is placed in position 378 and *lo hizo el pasado 1 de abril.* is marked backwards and overwritten with *despidió marzo*. *lo_* is marked, the space between *por debajo de_lo que* is deleted, the *o* in *lo que* is then deleted before *_nivel en el* is written.	2 contextual revisions	1/3	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.[10] → 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**.[11] → Mañana, cuando las plazas del Viejo Continente retomen su actividad, el selectivo español arrancará la sesión 1.000 puntos por debajo del **nivel en el que despidió marzo**.[12]
Pause: 3089 ms						

The extract above illustrates the temporal sequence of two revisions made to the third and last sentence of the article at the time. After a pause of about 15 seconds during which, according to the observation protocol and the retrospective interview, the author uses her personal calendar to clarify her doubt about a date, in the first revision, R1, she marks the last part of

[10] 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.

[11] 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.

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

the sentence, *lo hizo el pasado 1 de abril*, and overwrites it with *despidió marzo*. In the next step, R2, the author deletes the space between *de* and *lo que,* followed by the *o* in *lo que* before writing *nivel en el*. This takes place between 7 minutes 29 seconds and 7 minutes 46 seconds in the process.

3 Analysing online revisions

When tracking the text production process as it unfolds in computer-based writing, the continuous revisions made as part of the ongoing text production process become visible to the researcher. Because these online revisions are actions which are continuously shaping the text, certain issues must be considered when they are to be analysed and their effect on the text is to be interpreted. In the following, some of these issues will be addressed.

3.1 Categorising online revisions

At any given point during writing, the written text can be revised at its leading edge, also known as the *point of inscription* (e. g. Matsuhashi 1987, Stevenson/Schoonen/de Glopper 2006), where new text is being transcribed, and it can be revised in the text already written, i. e. after the text has been transcribed. When revisions occur at the leading edge, no textual context is transcribed following the revision, only before (cf. Example 1 below). Consequently, these revisions are pre-contextual, a term coined by Lindgren (2005). At the leading edge, the writer continually makes decisions about content and form while creating the emergent text (Matsuhashi 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). As opposed to revisions at the leading edge, revisions in the text already written are undertaken when writers move away from the leading edge to insert new text or to omit, substitute or rearrange already transcribed text (cf. Example 2 below). In this sense, these revisions are both preceded and followed by text. Thus, writers are operating within an already transcribed context, which makes the revisions contextual (Lindgren/Sullivan 2006b: 171).[13] The difference between the two revision categories is illustrated in the following two examples, which include data from one of the writing processes in the study.

In Example 1, the writer (P3) was working on the first sentence of her article (R0). As soon as she had transcribed the preposition *en* ('in') at the leading edge, she replaced it by another preposition, *desde* ('since'), while also transcribing the complement in the preposition phrase, *noviembre de 2010* ('November 2010') (R1).

[13] It appears from the text that Lindgren (2005) and Lindgren/Sullivan (2006b) use the two adjectives *pre-contextual* and *contextual* to refer to the difference in linguistic environment of a revision. In contrast, other scholars, e. g. from the Systemic-Functional tradition, use the term *co-text* to refer to the linguistic environment of a word, reserving *context* to mean "the context of language, a connotative semiotic system" (Matthiessen/Teruya/Lam 2010). However, as this article takes its point of departure in the framework of Lindgren/Sullivan (2006b), it makes most sense to adhere to their terminology.

(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

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

Example 1

When this revision was being made, it constituted the leading edge of the transcribed text, i. e. there was only transcribed text preceding the revision and no textual context transcribed following the revision. Accordingly, the revision is pre-contextual.

In Example 2, the text producer (P3) had completed the 9th sentence (R0). After a short pause of barely 2.5 seconds, the journalist, hitting *Enter*, created space for a new paragraph, which she initiated after another 2 seconds, then stopped transcribing the second word. After a little more than 4 seconds, the author returned to the beginning of the previous sentence, marking *el peso* ('the weight') and overwriting it with *la proporción* ('the proportion') (R1).

(R0)
Así, el peso que los inversores no residentes tienen en el reparto de la tarta de deuda soberana ha pasado del 50,4 % al 37,5 % en tan sólo tres meses.
Thus, the weight that non-resident investors have in the distribution of the sovereign debt pie has gone from 50.4 % to 37.5 % in just three months.

(R1)
Así, la proporción que los inversores no residentes tienen en el reparto de la tarta de deuda soberana ha pasado del 50,4 % al 37,5 % en tan sólo tres meses.
Thus, the proportion that non-resident investors have in the distribution of the sovereign debt pie has gone from 50.4 % to 37.5 % in just three months.

Example 2

Then the cursor was moved back to the leading edge of the text, and the transcribing of the word was resumed. The revision was made within an already transcribed context preceded and followed by text, hence this is a contextual revision.

3.2 Interpreting the effect of online revisions

The distinction between pre-contextual and contextual revisions according to their location, i. e. in the text already transcribed (*contextual revision*) or in the text currently being transcribed (*pre-contextual revision*), is relevant when the effect of a revision 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 is intending 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, that text being the only one, which the researcher can safely assume the writer to be conscious of when making the revision. Accordingly, pre-contextual revisions lack analytical context, which makes their effect on the text difficult – even impossible – to interpret on the basis of an observation of the writing process, e. g. by keystroke logging. This is excellently illustrated by the Example 1 in Section 3.1 above. In this example, the prepositions, *en* ('in') and *desde* ('since'), are contingent on the following text, which is yet to be transcribed at the time of the revision. Hence, the revision is not made in a semantically meaningful context that may qualify an interpretation as to whether the revision reflects a change in the non-transcribed content, or whether it affects the form of the text (see Section 3.3 regarding the concept of semantically meaningful context and Section 3.4 regarding the content-form dichotomy). By contrast, in Example 2 in Section 3.1 above, the substitution of the noun, *peso* ('weight'), by another, *proporción* ('proportion'), is made within a semantically meaningful context which allows for the interpretation of the revision as only involving the form of the text, leaving the content unaffected. As appears from the above, only the effect of revisions made in a semantically meaningful context is interpretable on the basis of keystroke logging alone. The interpretation of the effect of revisions made in semantically non-meaningful contexts often – if not always – requires a verbal protocol from the writer, e. g. *think aloud* or *stimulated recall,*[14] as a supplement. Hence, such revisions were not included in the study.

3.3 Analytical framework

The approach to the analysis and interpretation of revisions was inspired by the online revision taxonomy developed by Lindgren and Sullivan (2006a, 2006b) in collaboration with Marie Stevenson (Stevenson/Schoonen/de Glopper 2006). The taxonomy categorises textual revisions both according to their location, e. g. pre-contextual or contextual revisions, and to their effect on the text, i. e. revisions of content or form (cf. Section 3.4 below). However, due to the distinguishing features of the two revision categories, the taxonomy proved to be insufficiently accurate to be operationalised, and too coarse to categorise all interpretable revisions in the data (see Haugaard 2016 for an in-depth discussion of the taxonomy and its shortcomings). For the purpose of analysing all interpretable revisions in the data, a stringent and nuanced analytical framework was developed based on a heuristic and highly iterative analysis characterised by interaction between the established theory and the data. This new analytical framework introduced the concept of semantically meaningful context, i. e. a group of words that are syntactically and semantically linked, and whose content constitutes a potentially complete whole[15] which makes sense in itself. It was suggested to categorise online revisions made during text production on a continuum of semantically meaningful context. At the one end of the continuum lies the potentially most complete semantically meaningful context represented by a sentence concluded by a sentence-completing character, e. g. a full stop or a question mark. At the other end, the semantically non-meaningful context is placed, represented by contexts lacking

[14] *Think aloud* is a research method whereby a writer verbalises his or her thoughts during text production. The verbalisations are often recorded and transcribed into a protocol. *Stimulated recall* is a research method whereby a writer is invited to recall his or her concurrent thinking during a writing episode when prompted by some form of visual recall.

[15] The concept of potentiality accounts for the fact that no context can be defined as completed until the article has been printed in the newspaper.

semantic meaning. These semantically non-meaningful contexts will often coincide with the leading edge of the text, but as writing on a computer is not necessarily a linear process, contexts lacking semantic meaning are also to be found when writers move back in the text already transcribed, for instance to add a new element, and during the addition of this element, make additional revisions before a new semantically meaningful context is transcribed. In between the two ends, the continuum holds semantically meaningful contexts that are potentially less complete, such as semantically meaningful sentences without sentence-completing characters and semantically meaningful phrases (see Haugaard 2016 for the reasoning and an exemplification of the different types of contexts). Contextual revisions are made in semantically meaningful contexts – some even *of* semantically meaningful contexts, such as the substitution of one sentence with another. By contrast, pre-contextual revisions are characterised by the fact that they occur in semantically non-meaningful contexts and are located to the far right of the continuum. With the exception of very few revisions in semantically non-meaningful contexts, including a particular group involving the retranscription of (parts of) the just deleted text (cf. Example 3 below), only the effect of revisions made in semantically meaningful contexts can be interpreted solely on the basis of keystroke logging. Accordingly, these are the only revisions interpreted in the study. Retranscription of (parts of) the just deleted text occurs when deleting text is used as a means to reach the place in the already transcribed text where the revision is to be carried out. These originally pre-contextual revisions, typically at the leading edge of the text, are used to omit or substitute previously (partially) written text and to add new text, as in the following example where the noun *mañana* ('tomorrow') is added.

(R0)
En los últimos 30 días, el Ibex 35 ha perdido 1.000 puntos y arra
In the last 30 days, the Ibex 35 has lost 1,000 points and [arra?]

(R1)
En los últimos 30 días, el Ibex 35 ha perdido 1.000 puntos y
In the last 30 days, the Ibex 35 has lost 1,000 points and

(R2)
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, the Ibex 35 has lost 1,000 points and tomorrow will start the month of May

Example 3

By introducing, at this level of analysis, the interpretation as to whether the revision is conducted in a semantically meaningful context, the analytical framework distances itself from a more objective categorisation of the location of revisions at the leading edge or in the transcribed text. This allows for a systematisation of the contexts in which the effect of revisions at the leading edge can be interpreted and the contexts in which the effect of revisions made in already transcribed text cannot be interpreted.

3.4 The content-form dichotomy

To operationalise the content-form dichotomy, the analysis was based on Faigley/Witte's (1981) understanding of the two concepts. According to Faigley/Witte, content is not to be

understood restrictively as concepts explicitly referred to in the extant text, but also as concepts which can be reasonably inferred from it. In order to anticipate that inferences may vary from reader to reader, the focus is on the inferences which the writer "raises to the surface by adding explicit text or requires by deleting explicit text during revision" (Faigley/Witte 1981: 402). Accordingly, the dichotomy is based on whether the writer makes content explicit so that readers do not need to infer relations from the text, or whether the writer fails to explicate and leaves it to the reader to make inferences. Faigley/Witte (1981: 402) illustrate their argument with the following short text:

1. I just made it to the station on time.
1A. I got on the train.
2. I had to buy my ticket from the conductor.

They state that the reader will be able to infer sentence 1A if it is omitted. That is, the reader is able to make the inference about the information *I got on the train*. Hence, the omission or addition of sentence 1A will not affect the content of the text, and, as a result, this type of revision is classified as revision of form, leaving the content untouched. Accordingly, revision of form involves changes that rephrase the content without altering it. Beside this distinction between content and form, Faigley/Witte (1981: 403) also distinguish six different revision types, namely *addition*, *deletion*[16], *substitution*, *permutation*, *distribution* and *consolidation*, which apply in relation to revisions of content and form alike.

Summing up, revisions of form neither omit nor substitute original content that cannot be inferred from the written text as it is, nor do they add content that cannot already be inferred. By contrast, revisions that affect the content of the text add new content or omit existing content that cannot otherwise be inferred from the written text.

The analysis of the revisions and particularly the interpretation of their effect on the text were carried out several times. These processes were continuously and thoroughly discussed with experts, and a number of revisions have been the subject of repeated and nuanced discussions.

4 Overview of results

The following overview of results focuses on the overall features of the composing processes of the three articles in terms of revision practices.

[16] As a revision type, *deletion* should not be confused with *omission*. *Omission* of text will always imply the activity of *deleting* the text in question, but *deletion* of text does not always entail the *omission* of the text in question. Consequently, I will use the term *omission* with reference to the activity whereby text is deleted and omitted, the reader being forced to infer what was previously explicit.

Table 3: Overview of revision type and effect in the overall writing process

Revision type	Revisions during the ongoing text production						No. of revisions	Revisions during the systematic review						No. of revisions	Total no. of revisions
	Content			Form				Content			Form				
	P1	P2	P3	P1	P2	P3		P1	P2	P3	P1	P2	P3		
Addition	7	4	11	12	9	6	**49**	1						**1**	**50**
Substitution	1	2	6	8	6	21	**44**				7		9	**16**	**60**
Omission				2	2	1	**5**			1	5		9	**15**	**20**
Distribution					1	1	**2**								**2**
Total (participant)	8	6	17	22	18	29	**100**	1		1	12		18	**32**	**132**
Total (effect)	**31**			**69**			**100**	**2**			**30**			**32**	**132**
Total	**100**							**32**							**132**

The analysis of revisions during the writing of the three news articles showed that their number differed from process to process. Thus, during their work with the texts, and as shown in Table 4 below, P1 carried out a total of 43 revisions, P2 24 revisions in total, and P3 65 revisions. When these numbers of revisions are related to the number of characters typed during the processes (cf. Table 1), it appears that P1 and P3 carried out roughly the same number of revisions per 100 words produced in the articles, i. e. P1 made 8.1 revisions per 100 words, and P3 made 8.7 revisions per 100 words. By contrast, P2 only carried out 3.7 revisions per 100 words, which are significantly fewer revisions per 100 words produced. This heterogeneity will be accentuated in Section 4.1 below and further discussed in Section 5.

Table 4: Revisions related to produced characters and according to time of occurrence

Participant	Total number of revisions during the writing process	No. of revisions related to produced characters in total	No. of revisions during ongoing text production	No. of revisions during systematic review	No. of revisions related to produced characters during ongoing text production
P1	43	8.1	30 (69.8 %)	13 (30.2 %)	5.6
P2	24	3.7	24 (100 %)	0 (0 %)	3.7
P3	65	8.7	46 (70.8 %)	19 (29.2 %)	6.2

4.1 Time of occurrence

Heterogeneity was also demonstrated in the time of occurrence of the revisions. The analysis of the revision types and their effect on the text showed that revisions were distributed between what seems to be two different phases of the writing process, i. e. ongoing text production, during which cohesive and coherent text for the article was produced, and a systematic review of the potentially finalised text, during which the text was evaluated and its volume reduced. These findings will be further explained in Sections 4.2.1 and 4.2.2. However, as appears from

Tables 3 and 4 above, P2 carried out her writing as one phase; hence, she did not carry out a final systematic review of the potentially final text, which stands in contrast to both P1 and P3. This circumstance may help to explain P2's rather low number of revisions. Nonetheless, even if this systematic review of the text is ignored, P3 and P1 still carried out more revisions per 100 words produced than P2, i. e. 6.2 revisions and 5.6 revisions per 100 words produced, respectively, against P2's 3.7 revisions per 100 words produced. This difference in number of revisions cannot be explained on the basis of the study. However, see Section 5 for a discussion of feasible causes and possible further investigations.

As shown in Tables 3 and 4 above, P1 and P3 both revised considerably more frequently during the ongoing text production than during the systematic review. Thus, P1 carried out 30 revisions in the ongoing text production, and only 13 revisions during the systematic review, which corresponds to 69.8 % and 30.2 %, respectively. P3 revised 46 times during the ongoing text production, compared to 19 times during the systematic review, which corresponds to 70.8 % and 29.2 %, respectively. Accordingly, P1's and P3's revisions were distributed fairly evenly between ongoing text production and systematic review.

4.2 Distribution of revisions across content and form

The analysis of the way in which the revisions were distributed across content and form also demonstrated a homogeneous picture of the three writing processes. As is shown in Table 3 above and Figure 1 below, all three journalists made significantly more form revisions than content revisions. P1's 34 revisions of form account for 79.1 % of her 43 revisions, and only nine of her revisions, or 20.9 %, change the content. P2's 18 form revisions amount to 75 % of her 24 revisions, and the remaining 25 %, or six revisions, affect the content. P3's 47 revisions of form correspond to 72.3 % of her 65 revisions, and her 18 revisions of content correspond to 27.7 %.

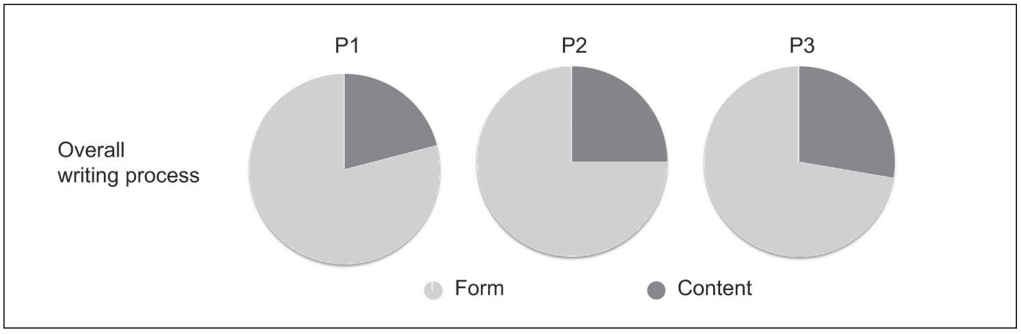

Figure 1: Journalists' revisions distributed across content and form in the overall writing process

Consequently, P2 and P3 both revised the form of their texts three times or approximately three times as frequently as they revised the content, and P1's revisions were related to form approximately four times as often as to content.

4.2.1 Content revisions

During writing, all three journalists added and substituted content (cf. Table 3 and Figure 2). Being the only journalist to do so, P3 also omitted content. One major similarity between P1 and P3 in terms of content revisions is that they both revised content significantly more often during the ongoing text production than during the systematic review of the potentially final text; they both undertook just a single content revision in their systematic review of the text (cf. Table 3). In other words, 88.9 % of P1's revisions of content and 94.4 % of P3's were undertaken during the ongoing text production and only 11.1 % and 5.6 %, respectively, during their final review of the text.

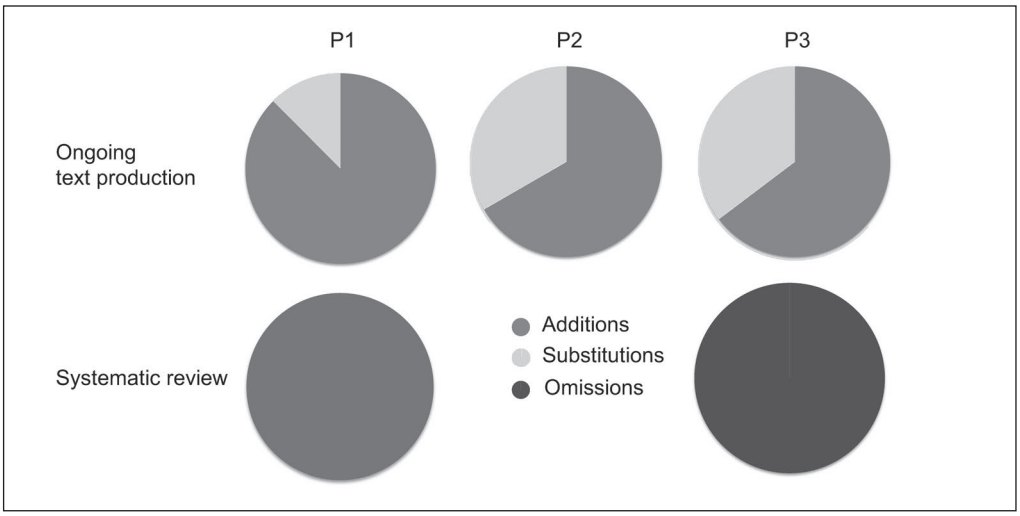

Figure 2: Journalists' content revision types

As illustrated by Figure 2 above, during the ongoing text production, all three journalists' content revisions were characterised by adding or substituting content (cf. also Table 3), by far the most common revision type being adding. Content additions added new content, either in an existing sentence or as a new sentence. In line with all of the substitutions, one content addition in an existing sentence appeared to correct a content error.

During the systematic review of the text, P1 added a sentence, which appeared as sentence 23 (of 28) in the printed article, whereas P3 omitted the two last sentences of the potentially completed article[17] (cf. Figure 2 and Table 3). Accordingly, content revisions during the systematic review involved entire sentences.

In summary, content was not omitted during the ongoing text production, nor was it substituted during the systematic review. This indicates that the focus of the first phase was on producing suitable content to the article by adding and substituting content, but not omitting it. Moreover, it suggests that the focus of the second phase was on adding missing content and omitting unnecessary content, but not on changing content by substituting it.

[17] Even though the revision involves the deletion of two sentences, it represents one revision because it is carried out as one single activity.

4.2.2 Form revisions

Working with the form of their texts, all three journalists added, omitted and substituted text, although with different frequencies (cf. Table 3 and Figure 3). Moreover, P2 and P3 amended the distribution of text once each.[18] As was the case with content revisions, P1 and P3 both revised the form of their text more often during the ongoing text production than in their systematic review of the potentially final text; 64.7 % of P1's form revisions and 64.1 % of P3's were undertaken during the ongoing text production, and only 35.3 % and 38.3 %, respectively, in their review of the potentially final text.

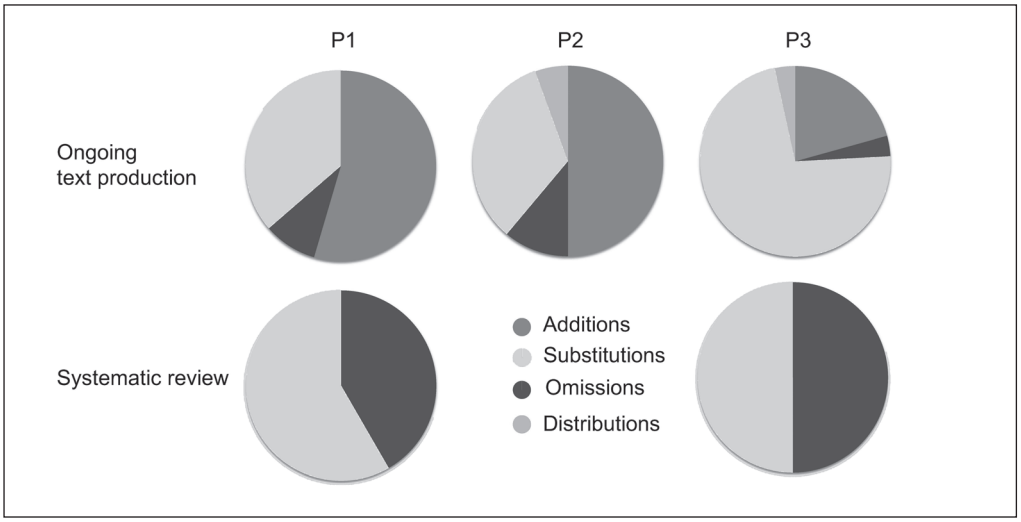

Figure 3: Journalists' form revision types

As illustrated by Figure 3 and Table 3 above, during ongoing text production, the form revisions of all three journalists were characterised by adding, substituting and omitting text, albeit in a varying degree. Furthermore, P2 and P3 distributed text once each. During this phase, the two most common revision types were adding and substituting. A number of additions of form were characterised by making explicit what could already be inferred from the text. Moreover, additions and substitutions of form repaired grammatical and orthographic errors such as missing characters, quotation marks and commas.

During their systematic review of their text, P1's and P3's form revisions were characterised by the omission and substitution of text (cf. Figure 3 and Table 3). The majority of the substitutions reduced the volume of the text involved, either by substituting one phrase with a corresponding shorter phrase or by paraphrasing the original text segments, thus reducing their size.

Summing up, during ongoing text production the most frequent revision types were addition and substitution, and during the systematic review, the most frequent revision types consisted in omission and substitution. This suggests that the focus of the first phase was on

[18] According to Faigley/Witte (1981: 403), "distributions occur when material in one text segment is passed into more than one segment".

developing the content and making relations explicit. Moreover, the high number of revisions which omitted text during the systematic review indicates that a focus in this phase was on shortening the article. This is supported by the circumstance that the majority of revisions in which text was substituted contributed to reducing the volume of the article.

5 Conclusion and discussion

From the above section it appears that the overall result of the study presented a relatively homogenous picture, including certain variations, in which the form of the text was revised markedly more often than the content (approx. three to four times more often). Thus, the initial heterogeneity in terms of the participants' years of employment at the business section and their expertise within the subject areas did not seem to be reflected in these overall results. The uneven distribution of revisions between content and form suggests that the journalists were skilled at planning and generating suitable content which did not often require revision, which may be a result of the highly standardised format of a news story (see also below) and the routinised writing processes resulting from several years of experience: The structure of the news story and the content of the different subject areas may be so well-known to the journalist that more capacity was available for shaping the form of their news stories.

The overall results also seem to suggest that the revision types and their effect on the text reflect the diverging purposes of these two phases: The first phase served to generate cohesive and coherent text for the article, and the second phase aimed to evaluate and, in particular, to reduce the volume of the written text. Thus, these results do not support Perrin's findings (2001 referred in Perrin 2003: 919) that experienced text producers "are more likely to revise their texts in several, complete passes through and gain a certain distance in between" (Perrin 2003: 919). Perrin's conclusion is based on 17 case studies of journalists' text production at the workplace, often including several texts from the same journalist/workplace as well as a variety of genres. In this respect, it is important to emphasise that the study reported in this article explored revisions to the news article genre, which is a "highly structured genre [...] involving the 'inverted pyramid' form" (Schumacher/Scott/Klare et al. 1989: 392) as well as a very tight timeframe. As "journalistic writing involves genres of widely varying constraints" (Schumacher/Scott/Klare et al. 1989: 392), this circumstance may be of consequence to the observed revision practices in relation to other and less constrained journalistic genres such as editorials (Schumacher/Scott/Klare et al. 1989: 393), and, on a more general note, in relation to other types of professional text production with other characteristics and constraints. More studies are therefore needed to enhance our knowledge of the impact of specific genres and timeframes on text production and revision practices.

For the purpose of analysing the revisions in the study, a stringent and nuanced analytical framework was developed based on existing theories and the empirical material. Accordingly, the framework was tailored to analysing and interpreting online revisions during the production of authentic news texts. To what extent the framework applies in other contexts, future studies will show.

The study presented here explored revision practices in journalistic text production. However, revisions constitute only one aspect of a text producer's work on the content and form of the text. Pauses make up another and complementary aspect which may reflect evaluations and mental revisions (Lindgren/Sullivan 2006a: 38) and may thus influence the quality of the written text and also the number of textual revisions (Witte 1987: 401). As mentioned above,

P2 carried out fewer revisions per 100 words in the completed article than the other two participants, and it may therefore be assumed that she succeeded, to a greater extent, in creating quality in her mental text prior to transcription, entailing that the written text did not need revision as frequently as the other two texts. Hence, a study of the pauses would enrich the description of the journalists' revision practices and might contribute to an understanding of any differences in the frequency of revisions among the participants: Are longer and/or more frequent pauses a possible explanation for the number of textual revisions?

According to the theory, the interpretation of online revisions informed by keystroke logging should only relate to the revision's effect on the text, and not to the underlying reasons (e. g. Lindgren/Sullivan 2006b: 160). In the present study, however, it was found that surrounding activities (e. g. other revisions and mouse movements) as well as the written context might in certain cases lead to a deeper understanding of the effect of revisions beyond the distinction between content and form, and, thus, to an insight into to what (the effect of) the revision contributes. Accordingly, it appeared that in addition to adding new content to the text, the insertion of new sentences seemed to accentuate the coherence of the text. Similarly, additions of form that made explicit what could already be inferred from the text strengthened the cohesion of the text, helping the reader to draw inferences. Moreover, some substitutions of form appeared to indicate that linguistic expressions were related to one another, thus bringing about variation in the terminology, e. g. by using lexical anaphors. This observation indicates that the textual context and a comprehensive overview of the writing process might be fruitful in enhancing the understanding of how the text producers juggle content and form.

To sum up, the exploratory and qualitative nature of the study has provided a detailed analysis of the journalists' revision activities and has offered nuanced insights into their text production. In this sense, the study contributes to our empirical knowledge of how journalists write and revise news texts. Moreover, the study contributes to our theoretical-methodical knowledge of how to explore, analyse and interpret online revisions in authentic news texts. However, due to the limited scope of this study, further studies should be conducted in order to gain further insights into these aspects. Nonetheless, the overall tendency of the analysis and the details which it reflects may be used as a point of departure for new studies and can help generate hypotheses about how other text producers, both in similar and different contexts, write and revise theirs texts, and how they juggle content and form.

References

American Press Institute = https://www.americanpressinstitute.org/journalism-essentials/

Carl, Michael (2012): "Translog – II: a Program for Recording User Activity for Empirical Reading and Writing Research." *Proceedings of the Eighth International Conference on Language Resources and Evaluation,* European Language Resources Association (ELRA). *International Journal of Computational Linguistics and Applications (IJCLA)* 3.1: 153–162.

Creswell, John W. (2014): *Research Design. Qualitative, Quantitative, and Mixed Methods Approaches.* 4th edition (International Student Edition). Thousand Oaks, CA: Sage.

Dam-Jensen, Helle/Heine, Carmen (2009): "Process Research Methods and Their Application in the Didactics of Text Production and Translation." *Trans-kom* 2.1: 1–25. 27.04.2017 <http://www.trans-kom.eu/bd02nr01/trans-kom_02_01_01_Dam-Jensen_Heine_Process_Research.20090721.pdf>.

Faigley, Lester/Witte, Stephen (1981): "Analyzing Revision." *College Composition and Communication* 32.4: 400–414.

Hartley, Jannie Møller (2011): *Radikalisering af kampzonen. En analyse af netjournalistisk praksis og selvforståelse i spændingsfeltet mellem idealer og publikum*. PhD thesis, Roskilde University.

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.

Jakobs, Eva-Maria/Perrin, Daniel (2014): "Introduction: Theory and methodology in text production research." *Handbook of Writing and Text Production*. Eds. Eva-Maria Jakobs/Daniel Perrin. (Handbooks of Applied Linguistics/HAL 10). Berlin: De Gruyter Mouton. 27–30.

Johnson, Burke R./Onwuegbuzie, Anthony J./Turner, Lisa A. (2007): "Towards a definition of mixed methods research." *Journal of Mixed Methods Research* 1.2: 112–133.

Leijten, Marielle/Van Waes, Luuk (2013): "Keystroke Logging in Writing Research: Using Inputlog to Analyze and Visualize Writing Processes." *Written Communication* 30.3: 1–35.

Leijten, Marielle/Van Waes, Luuk/Schriver, Karen/Hayes, John R. (2014): "Writing in the workplace: Constructing documents using multiple digital sources." *Journal of Writing Research* 5.3: 285–337. 15.05.2017 <http://www.jowr.org/articles/vol5_3/JoWR_2014_vol5_nr3_Leijten_et_al.pdf>.

Lindgren, Eva (2005): *Writing and Revising. Didactic and Methodological Implications of Keystroke Logging*. PhD thesis, Umeå University.

Lindgren, Eva/Sullivan, Kirk P. H. (2006a): "Writing and the analysis of revision: An overview." Sullivan/Lindgren (2006): 31–44.

Lindgren, Eva/Sullivan, Kirk P. H. (2006b): "Analyzing on-line revision." Sullivan/Lindgren (2006): 157–188.

Matthiessen, Christian M. I. M./Teruya, Kazuhiro/Lam, Marvin (2010): *Key Terms in Systemic Functional Linguistics*. London: Continuum.

Matsuhashi, Ann (1987): "Revising the text and altering the plan." *Writing in real time. Modeling text production processes*. Ed. Ann Matsuhashi. Norwood, NJ: Albex. 197–223.

Neergaard, Helle (2007): *Udvælgelse af cases – I kvalitative undersøgelser.* 2nd edition. Frederiksberg C: Samfundslitteratur.

Perrin, Daniel (2001): *Wie Journalisten schreiben*. Konstanz: UVK.

Perrin, Daniel (2003): "Progression analysis (PA): investigating writing strategies at the workplace." *Journal of Pragmatics* 35: 907–921.

Perrin, Daniel (2013): *The Linguistics of Newswriting*. (AILA Applied Linguistics Series 11). Amsterdam/Philadelphia: Benjamins.

Perrin, Daniel/Ehrensberger-Dow, Maureen (2006): "Journalists' Language Awareness: Inferences from Writing Strategies." *Revista Alicantina de Estudios Ingleses* 19: 319–343.

Pitts, Beverley J. (1982): "Protocol analysis of the newswriting process." *Newspaper Research Journal* 4: 12–21.

Schultz, Katherine (2006): "Qualitative Research on Writing." *Handbook of Writing Research*. Eds. Charles A. MacArthur/Steve Graham/Jill Fitzgerald. New York/London: Guilford. 357–373.

Schumacher, Gary M./Scott, Byron T./Klare, George R./Cronin, Frank C./Lambert, Donald A. (1989): "Cognitive Processes in Journalistic Genres. Extending Writing Models." *Written Communication*. 6.3: 390–407.

Stake, Robert E. (1995): *The art of case study research*. Thousand Oaks, CA: Sage.

Stevenson, Marie/Schoonen, Rob/de Glopper, Kees (2006): "Revising in two languages: A multi-dimensional comparison of online writing revisions in L1 and FL." *Journal of Second Language Writing* 15: 201–233.

Sullivan, Kirk P. H./Lindgren, Eva, eds. (2006): *Computer Keystroke Logging and Writing: Methods and Applications* (Studies in Writing 18). Oxford: Elsevier.

Thagaard, Tove (2012): *Systematik og indlevelse – en indføring i kvalitativ metode*. Translated by Ole Lindegård Henriksen. København: Akademisk Forlag.

Van Hout, Tom (2010): *Writing from Sources: Ethnographic Insights into Business News Production*. PhD thesis. Department of Language and Communication, Ghent University.

Willig, Ida (2011): *Bag nyhederne – værdier, idealer og praksis.* 2nd edition. Frederiksberg C: Samfundslitteratur.

Witte, Stephen (1987): Pre-Text and Composing. *College Composition and Communication* 38.4: 397–425.

Yin, Robert K. (2003): *Case study research: Design and methods.* 3rd edition. Thousand Oaks, CA: Sage.

Dr. Rikke Hartmann Haugaard
Assistant professor in Spanish
professional and organisational communication
Aalborg University,
Department of Culture and Global Studies
Krogstraede 3, room 5.211
9220 Aalborg E
Denmark
haugaard@cgs.aau.dk
+45 9940 2799 / +45 27488779