

Modelling scientific certainty

Argumentation strategies vs. linguistic markers analysis

Luigi Pastore¹, Sara Dellantonio²

¹ Università degli Studi di Bari, Italy; luigi.pastore@uniba.it

² Università degli Studi di Trento, Italy; sara.dellantonio@unitn.it

Abstract. In recent years, there has been increasing interest in investigating science communication. Some studies that address this issue attempt to develop a model to determine the level of confidence that an author or a scientific community has at a given time towards a theory or a group of theories. A well-established approach suggests that, in order to determine the level of certainty authors have with regard to the statements they make, one can identify specific lexical and morphosyntactical markers which indicate their epistemic attitudes. This method is considered particularly appealing because it permits the development of an *algorithmic model* based on the quantitative analysis of the occurrence of these markers to assess (almost) *automatically* and *objectively* the opinion of an author or the predominant opinion of a scientific community on a topic at a given time. In this contribution we show that this line of research presents many kinds of problems especially when it is applied to research articles (rather than to popular science texts and basic research reports). To this aim, we propose two main lines of reasoning. The first one relies generally on the argumentative structure of scientific articles and shows that certainty/uncertainty markers are used differently in different argument forms and that therefore their number/frequency of use does not offer reliable indications for determining whether the topic at issue is considered by the authors to be more or less factual/speculative. The second one is based on the analysis of a sample of psychiatric research articles on homosexuality written over a long time span and taken from *The British Journal of Psychiatry*. Since the psychiatric perspective on homosexuality changed radically during the decades in which these articles were published, they offer an inventory of various kinds of argumentative strategies directed both at defending and confuting dominant as well as marginal positions. We focus especially on uncertainty markers and show that frequently the positions stated using expressions indicating uncertainty are actually not considered as conjectural or speculative by their authors, but that the use of uncertainty markers is motivated by a number of different and often incongruent rhetorical strategies.

Keywords: hedging, boosting, argumentation, certainty/uncertainty in science; scientific writings

1. Introduction

In recent years, there has been increasing interest in examining and modeling the pragmatic aspects of information exchange in real communication dynamics (Kamio 1997; Ifantidou 2005) as well as in investigating how these models might be applied to the analysis of oral and written science communication (Hyland 1995, 1996a, 1996b, 2002; Aguilar 2008; Heritage 2011). This work pursues a number of different aims: some research investigates scientific communication *per se* (see e.g. Crismore and Farnsworth 1990; Hyland 1997, 2011; Heritage 2011); while other research addresses the possible differences in communication across scientific disciplines (see e.g. Skelton 1997; Argawal and Yu 2010; Chao and Hu 2014; Afsar et al. 2014); still other research relies on the analysis of communication to gain new insights into scientific and/or nonscientific reasoning (see e.g. Origg and Sperber 2000; Wilson 2000, 2005; Durik et al. 2008; Winter et al. 2015), etc. However, some studies addressing scientific communication seek a different and possibly more ambitious objective, i.e. to find a (possibly automatic) method to distinguish between certain (i.e. factual) and uncertain (i.e. speculative) information in texts. The contention is that this method assures that information can be correctly detected and extracted from the source. The goal of this line of investigation is to develop a model to determine the level of confidence that an author or a scientific community (that expresses itself through corpora of scientific texts or oral communications) has at a given time towards a theory or a complex of theories.

One of the methods considered to be the most *objective* for assessing the degree of certainty/uncertainty regarding the various hypotheses embedded in a scientific text consists in analyzing the occurrences of specific lexical or morphosyntactical certainty or uncertainty markers used by the authors (which are also called respectively *boosters* and *hedges*; see e.g. Hyland 1996b, 2004). Indeed, this approach apparently offers both a quantitative and a qualitative measure of the degree of certainty/uncertainty conveyed by a scientific text, since it quantifies the number/frequency of marker use to determine what information is certain and what is uncertain. This method is considered particularly promising because it could lead to the development of an *algorithmic model* based on the analysis of the occurrence of specific linguistic markers to assess (almost) *automatically* and *objectively* the opinion of an author or the predominant opinion of a community on a topic or at least the general attitude of one or more authors towards an issue at a given time. Thus, this model could also be used to take practical decisions based on scientists' confidence with respect to specific views. Recently, this project has gained the attention of the research field known as *Natural Language Processing* (NLP) whose aim is to make Turing's dream come true, i.e. to develop computational systems capable of processing (summarizing, translating, interpreting, answering questions, etc.) natural languages. (see e.g. Auger and Roy 2008; Vincze et al. 2008; Morante and Daelemans 2009; Kim et al. 2009; Farkas et al. 2010; Choi et al. 2012).

In this contribution we aim to show that this line of research on scientific texts is not as promising as it seems and presents a number of relevant problems. Since research articles are the most important source of 'fresh', 'first hand' scientific knowledge and therefore are also the most interesting locus in which to carry out this analysis, we focus specifically on research articles and we show that this kind of linguistic analysis is not adequate to assess whether authors consider the hypotheses they address in their papers to be more or less reliable or uncertain. To support this position, we discuss the relationship between two types of analysis that can be carried out on scientific papers. Specifically, we compare those analyses that aim to assess the degree of certainty/uncertainty of specific positions expressed in scientific papers by quantifying hedges and boosters with those analyses that use the classical techniques of argument evaluation (adopted e.g. by informal logic and critical thinking). On the basis of this comparison, we propose two challenges to the two different lines of argumentation presented by researchers who support the use of boosters and hedges as a measure of certainty/uncertainty.

i. The first argument is theoretical and takes into account how hedges and boosters are used in various argument forms. Its aim is to show that in the case of texts with an argumentative structure the global number/frequency of occurrence of hedges and boosters does not indicate whether the author considers the matter s/he deals with – i.e. the conclusions s/he draws – to be factual/certain or to be speculative/uncertain (*Thesis1*).

ii. The second argument relies on the examination of a sample of scientific articles on psychiatry taken from *The British Journal of Psychiatry* (BJP). In particular, we focus on papers addressing a very controversial psychiatric issue that has undergone many conceptual revolutions from the beginning to the end of the 20th century – i.e. *homosexuality*. We examine these articles using both linguistic markers assessment and manual reconstructions of their argumentative structure. Then, we compare the outcome of these two kinds of analysis in order to show that both the number of markers and their kind (hedges vs. boosters) are irrelevant with respect to the issue of assessing the degree of confidence or of commitment authors have with respect to the hypotheses they discuss. On the basis of this analysis we conclude that it would be a mistake to trust linguistic markers of certainty and uncertainty to distinguish factual from speculative knowledge (*Thesis2*).

2. Linguistic means to express un/certainty and the dream of an automatic method for distinguishing factual from uncertain knowledge

Studies that aim at distinguishing between speculative (i.e. uncertain) knowledge and factual (certain) information on the basis of specific linguistic cues originate both historically and theoretically from two main lines of research. The first line arises from the anthropological and cross-linguistic investigations initiated by Boas (1938: 133) on the grammatical expressions used in various languages to specify the source of conveyed information and is centered in particular on the issues of *evidentiality* and *epistemicity* or *epistemic position(ing)*. The notion of *evidentiality* is used to describe the problem of determining the means used in different languages to specify what kind of evidence supports specific statements or even to specify whether there is any evidence for making a specific claim – i.e. whether someone saw something in the first person, or whether s/he heard this information from someone else, etc. (for an overview on these studies see e.g. Aikhenvald 2003a, 2003b). The related notion of *epistemic position(ing)* refers to the ways in which languages express a speaker's attitude regarding the reliability of sentence content (whether s/he believes it, knows it, doubts it, etc.); in other words: whether s/he considers it more or less likely or reliable.¹ A second line of investigation was initiated by Lakoff (1973: 471) and concerns *hedges*, i.e. – according to Lakoff's definition – “words whose meaning implicitly involves fuzziness – words whose job is to make things fuzzier or less fuzzy.” In spite of the fact – acknowledged later by Lakoff himself (see Lakoff 1987: 122) – that the hedging devices identified in his 1973 paper are inadequate, this work gave rise to extensive research on hedging strategies, i.e. on the linguistic markers used in various languages to express caution (uncertainty) regarding specific claims. The opposite linguistic strategy, which aims to confer a high degree of reliability to statements or to communicate that they are based on well-grounded knowledge, is called *boosting* while the expressions used for boosting are considered *boosters* (for some early literature on these two notions see e.g. Holmes 1983, 1984, 1990).

Despite their differing theoretical roots, these two lines of study can be equated in terms of their common research objective to investigate (a) *how* certainty and doubt are conveyed in scientific communication (i.e. what are the expressions or markers adopted to achieve this aim) and (b) *why* authors use these stylistic devices to ‘hedge’ or to ‘boost’ their claims (i.e. what is the exact function of these expressions in scientific discourse).

(a) The communication of a writer's confidence regarding the contents s/he presents is achieved using specific linguistic means, which serve as caveats for the reader and indicate whether a claim must be taken with caution or whether it should be considered to be reasonably reliable. Literature on scientific writing tried therefore to identify – i.e. to classify and list – these linguistic markers in order to study their function in scientific papers. The various studies rely on more or less detailed lists of markers, but they all include approximately the same (kinds of) items (markers). Hyland (1996b, 2004) is one of the authors that has

¹ For an overview on epistemicity and for a discussion of its affinity to evidentiality see e.g. Chafe and Nichols 1986; Bednarek 2006; Cornillie 2009. Even though most authors agree that these two notions are positively related to each other (that they overlap, or that epistemicity is included in/derived from evidentiality), the debate is not unanimous on this aspect: for a brief overview of the different ways in which the relationships between these two notions have been interpreted see e.g. Dendale and Tasmowski (2001).

addressed the issue of classifying hedges and boosters most systematically. He extracted from the available literature a set of linguistic devices used to express an authors' commitment to the truth of the claims s/he makes and organized them into specific groups according to their kind. To produce a list of these markers that is as complete as possible is necessary particularly for studies that aim at investigating large corpora of articles using automatic search (for a short list of the most important lexical markers see Hyland 2004: 188-189). Manual search allows one to identify hedging devices that were not included in the initial catalogue.

According to Hyland (1996b), in spite of the fact that the literature has focused mainly on *modal verbs* (e.g. 'could', 'may', 'might', 'would', 'should'), there are also a large number of other devices used in scientific articles to express certainty and uncertainty. Moreover, caution towards a claim is often expressed using *epistemic lexical verbs* like 'indicate', 'suggest', 'appear', 'propose'. These verbs can be further differentiated into three main groups: *speculative* (e.g. 'suggest', 'propose', 'believe', 'speculate', etc.); *inferential*² ('infer', 'indicate', 'imply', etc.); and *evidential*. So-called evidential verbs include verbs expressing *quotative evidence*, i.e. they specify the source of an information and the reliability of this source or the commitment of the quoted authors with respect to the truth of a claim (e.g. author X *suggested* that ...; or the authors Y and Z *speculate* that ... etc.); and verbs expressing *sensory evidence* ('sensory' is used here in a metaphorical sense to indicate verbs describing sensory experience, especially visual experience, like 'appear', 'seem', 'observe', 'attempt to gain insight' etc.). Other means to express un/certainty are *epistemic adjectives* ('un/likely', 'possible', 'probable' etc.) or *epistemic adverbs* ('apparently', 'probably', 'essentially', 'relatively', 'presumably', etc.). Also so-called '*downtoners*', *adverbs* which lower the effect of a verb (like 'quite', 'almost', 'usually', 'partially', 'rarely', 'virtually', 'probably', 'generally', 'approximately', 'somehow' etc.) function as hedges. Even though Hyland does not mention it, we can hypothesize that there are corresponding '*uptoners*', *adverbs* (like 'completely', 'always', 'exactly', etc.) that function as boosters (for a concise list of the devices that fall into the classes mentioned above see Hyland 2004: 188-189). Finally, linguistic means to hedge or boost specific claims include so-called *strategic hedges* which consist of more complex expressions like 'we do not know whether...', 'one cannot exclude ...', 'we are aware of the concerns expressed by ...' etc. Markers that are related to the structure of a sentence including interrogative or hypothetical forms also belong in this class.

We will refer to this general class of linguistic means to express doubt and certainty indiscriminately as hedges/boosters, or as hedging/boosting devices or as certainty/uncertainty markers. Furthermore, we will speak of lexical and morphosyntactical markers to designate respectively the words and the structural elements in a sentence (verbal tenses, moods, interrogative or hypothetical forms, etc.) that serve as linguistic means to express certainty and uncertainty. From Hyland's discussion it emerges that hedging in scientific writing is more important than boosting. This is not surprising. In fact, in science we often do not need to emphasize or to stress our certainty with respect to specific claims. To express factual statements that we consider to be reliable we can simply use the indicative present (see e.g. Hyland 1996a: 435). However, other authors point out that the use of boosting strategies in scientific literature must also be considered relevant since they allow us to identify information which is explicitly presented as certain (see e.g. Rubin et al. 2006).

(b) The study of the specific linguistic strategies people use to express certainty and uncertainty has important applications in the field of academic writing for a number of different reasons. On the one hand, they might help in understanding the rhetoric of scientific writings and the motivations that drive the authors to weaken or enforce their commitment to the truth of the hypotheses they discuss. On the other hand, they are also interesting to understand how science actually works (see e.g. Hyland 1996b: 252). As the literature of the rhetoric of science makes clear, hedging strategies in particular are used for a number of different reasons (for an overview see e.g. Hyland 2012). Sometimes they are used to *persuade*

² Actually Hyland (1996b) calls this verb class 'deductive', but we prefer not to use this term since we consider it to be incorrect for the kind of verbs that he has in mind. In fact, he specifies that the verbs included in this class are related to *inferential* reasoning, while in another article he admits that science is mainly inductive: "In fact, my data indicates that few knowledge claims are presented in unmitigated form: induction and inference rather than deduction and causality characterize most arguments in scientific discourse" (Hyland 1996a: 435).

colleagues of the strength of their position and at the same time they aim to express their positions in a *cautious* and *modest* way, in the awareness that the search for truth in science is never definitive or conclusively confirmed. In this sense, hedging strategies can also be a means to show *accuracy* and *precision* (Hyland 1996b; Skelton 1997). Sometimes authors use hedging strategies to obtain the *consensus* of the community, i.e. to explicitly comply with certain views (Hyland 2012: chap. 9). In other cases, hedges accomplish other stylistic functions like being *polite*; showing deference towards the scientific community; they are also used as a means of *captatio benevolentiae* (see e.g. Brown, Levinson [1978](1987); Hyland 2005) or to avoid personal commitment (Coates 1983; Chafe and Nichols 1986; Liddicoat 2005).

To these observations on the general functions accomplished by hedging (and to some extent boosting) strategies, it must be added that their use varies depending on many factors; among these at least the subjective style of the writer and the disciplinary field of the paper must be considered particularly relevant (see e.g. Crismore et al. 1993; Hyland 1999, 2009). As for this second aspect, a number of studies carried out on corpora belonging to various fields state that there are consistent differences in the use of hedges and boosters made by different disciplines: in general, texts in the humanities tend to use these linguistic markers more liberally than those on more technical subject matters (see e.g. Coffin et al. 2003; Hyland 2009). Other authors instead call attention to the composition of research articles and specifically to their organization into sections (Introduction, Methods, Results, Discussion, Conclusion) arguing that the use of linguistic markers and more generally the style adopted in the different sections also varies widely, because it aims to accomplish different rhetorical functions (Lemke 1998; Hyland 2006).

However, while the studies cited above limit the scope of their research to the study of the rhetorical functions accomplished by linguistic markers for certainty and uncertainty in scientific writings, there are also different kinds of analysis addressing the usages of hedges and boosters that do not focus on the style or on the rhetoric of scientific writings *per se*, but rather pursue a different and much more ambitious project. This line of study starts from the assumption that certainty and (to a greater extent) uncertainty are conveyed using specific markers. Thus, hedges and boosters can help us identify which scientific hypotheses/views an author considers to be less reliable or more uncertain. As is suggested e.g. by Auger and Roy “From a linguistic point of view, the identification and automatic tagging of expressions of certainty/uncertainty in textual data is a *sine qua non* condition to enable the empirical study and modeling of how humans assess certainty through their use of language” (Auger and Roy 2008: 1860).

Furthermore, if we could develop particularly sophisticated methods of search and algorithms (possibly machine-learning algorithms) which are able to identify all and only the markers that actually indicate uncertainty, ruling out ambiguities and false positive results³, the distinction between certain and uncertain claims could be carried out automatically also on very large corpora of literature. This kind of natural language processing is for example the goal Agarwal and Yu set for their work: “Hedging is frequently used in both the biological literature and clinical notes to denote uncertainty or speculation. It is important for text-mining applications to detect hedge cues and their scope; otherwise, uncertain events are incorrectly identified as factual events. However, due to the complexity of language, identifying hedge cues and their scope in a sentence is not a trivial task. Our objective was to develop an algorithm that would automatically detect hedge cues and their scope in biomedical literature” (Agarwal and Yu 2010: 953). A very similar project is sketched by a number of authors who pursue the NLP program: “Identifying hedged information in biomedical literature is an important subtask in information extraction because it would be misleading to extract speculative information as factual information. In this paper we present a machine learning system that finds the scope of hedge cues in biomedical texts” (Morante and Daelemans 2009: 28). And further: “The CoNLL-2010 Shared Task was dedicated to the detection of uncertainty cues and their linguistic scope

³ As examples of possible false results obtained identifying uncertain statements merely on the basis of uncertainty markers Agarwal and Yu (2010: 954) cite sentences like: “We can now study regulatory regions and functional domains of the protein in the context of a true erythroid environment, experiments that have not been possible heretofore.” In addition, they point out that – in the case of complex sentences in which only one part/aspect is qualified as uncertain through the association with a marker like “Right middle and [probable right lower] lobe pneumonia” – we need to be able to distinguish between the certain and the uncertain information (indicated in the example using the square brackets).

in natural language texts. The motivation behind this task was that distinguishing factual and uncertain information in texts is of essential importance in information extraction” (Farkas et al 2010: 1).

This generic reference to ‘literature’ or ‘texts’ as the locus where this kind of research can be carried out offers a clue to understanding what the final ambition or the ultimate goal of this kind of research might be: the conversion of the subjective certainty/uncertainty expressed by single authors into a general form of certainty/uncertainty. In fact, if science is interpreted, as Kuhn’s sociological account (1970) suggests, as a collective and democratic enterprise, then the truth of a hypothesis is decided collectively depending on whether most scientists agree with it. If so, since scientific literature is the means researchers use to express their views and their results, hedges and boosters could help us determining the degree of reliability attributed by a scientific community to a specific hypothesis at a certain time. We could indeed check large corpora of scientific literature in a specific field and ‘calculate’ how many (i.e. what percentage of) researchers consider a certain hypothesis to be un/certain. If this were possible, the project of natural language processing would have achieved a great success, providing society with a method for determining what scientific results are trusted by the most scientists in a given time and thus what scientific results should be trusted by most by everyone at that time.

3. Scientific articles as arguments

A research program that aims at identifying an automatic method based on lexical and morphosyntactical markers to determine whether a position expressed in a single constituent of text (one sentence, or one part of a sentence or a small cluster of sentences) is considered to be (more or less) reliable (factual) or uncertain (speculative) is not unproblematic. In fact, positions like the one we mentioned in the previous section according to which hedges and (to a lesser extent) boosters are highly variable rhetorical means that accomplish many functions and that “can only be understood in terms of a detailed characterization of the institutional, professional and linguistic contexts in which they are employed” (Hyland 1996a: 434) already represent a criticism of these views and a possibly insurmountable obstacle to their fulfilment as they argue that hedging and boosting strategies are *not* used primarily to express certainty or uncertainty (we will go back to this below).

However, in order to consider the very possibility of pursuing this kind of project, let us assume, for the sake of argument, that hedges and boosters are indeed used at least mostly to express un/certainty. Even in this case – this is the point we will argue for in this section (*Thesis1*) – the project of using the kind and number of linguistic markers included in a text to determine whether the author was more or less certain of the positions expressed cannot be successful. The reason for this must be traced back to the argumentative structure of scientific writings and to their argumentative character as opposed to the descriptive nature of more simple kinds of writings like e.g. brief research reports or popular science which merely communicate observations carried out by the researchers.

Scientific articles do not present mere expositions of facts, but they are written in order to support or to challenge certain hypotheses. More precisely, they are long and complex arguments whose aim is to assess what views should be sustained or dismissed. Arguments consist in a number of statements logically related to each, consisting respectively in premises and conclusions. Their purpose is to give reasons for a conclusion. Sometimes they are used to justify or to refute a point of view on the basis of reasons, at other times they are used to offer explanations for something in terms of causes.⁴ We have deductive explanations in which the conclusion (the *explanandum*) is derived from some statements (the *explanans*) describing general principles or laws and the specific condition concerning the phenomenon to explain⁵; we have inductive explanations or inferences to the best explanation, in which we infer possible causes from a

⁴ Even though reasons and causes are clearly not one and the same thing, from the point of view of the logical structure of the argument we can consider them as equal (see Sinott-Armstrong and Fogelin 2009: 3-16).

⁵ According to the *Deductive Nomologic Model* developed by Carl Gustav Hempel and Paul Hoppenheim this is the only appropriate way to develop scientific explanations. See Hempel and Hoppenheim 1948; Hempel 1965; Ladyman 2002.

number of situations⁶. Reasons and causes might also be combined in the same argument when we need both to individuate causes that make something happen and reasons to justify why certain hypotheses involving the causes are stronger than others.

The complex arguments scientific articles consist of usually start from observational data and/or hypotheses accepted in the literature which they use as premises to infer the conclusions. Some intermediate conclusions are then used in connection with some supplementary hypotheses as additional premises to draw further conclusions. Confirmation is the simplest form of argumentation: an author has some evidence that s/he considers more or less certain and s/he uses this evidence to support a certain conclusion. In the case of confirmation, the certainty of the conclusion depends on the certainty of the premises. For this reason, we might expect that a high number of hedges in the premises will indicate that the conclusion is uncertain and that the topic considered is basically speculative. Conversely, it can also be assumed that a high number of boosters (or at least the absence of hedges) will show that the conclusion is certain and that the topic is generally considered factual.

However, scientific research is rarely only about confirmation. Scientific inquiries frequently aim to refute some well-established view or to argue for a position through the refutation of alternative possibilities. Most often, scientific writings consist of complex arguments including different augmentative forms (refutation; confirmation through linked or convergent premises, a chain of inferences, etc.). All of these argumentative forms differ in the way the premises provide support for the conclusion and one cannot take for granted that the conclusion is asserted with a high degree of certainty only when the premises are considered to be certain as well or, conversely, that the conclusion is asserted with a low degree of certainty only when the premises are considered to be uncertain. This means that the number and the quality of hedging or of boosting devices used in an argument to characterize the statements that are used to describe the premises as more or less reliable are not necessarily predictive of the strength (the degree of certainty/uncertainty) with which the conclusion is asserted starting from these premises.

Let us consider several cases more closely.

(I) We could have e.g. an argument in which the premises are doubtful, and therefore the conclusion must be rejected. In this situation, the author might use a high number of hedges in the premises to indicate their implausibility. However, the argument can be strong: the author can be very sure that – because of the weakness of the premises – the conclusion must be rejected. Thus, globally we will have a high number/frequency of hedges, even though the argument and its conclusion are considered as strong.

(II) A similar but opposite situation can occur when the link between premises and conclusion is weak, i.e. when the premises offer only indirect evidence for the conclusion. Here the premises might be presented as certain (using corresponding markers), but the conclusion can still be very uncertain because of its loose link to the premises. In this case, even though the global number/frequency of hedges is extremely low and/or the number/frequency of boosters is high, the author might be uncertain of his/her conclusion.

(III) Analogously, in the case of linked or convergent arguments (arguments in which more premises are combined in order to support a conclusion or in which more premises provide independent reasons for supporting the conclusion), it is possible that several uncertain premises – presented as such using corresponding markers – give rise to a fairly strong argument because taken together they offer strong support for the conclusion. Also in this case, the global number of hedging/boosting devices is not predictive of the strength of the argument or indicative of the certainty of the author on the matter.

Because in all these cases there is no correspondence between the kind/number of linguistic markers and the actual strength of the argument, these argumentative strategies represent a serious challenge for the project of identifying an automatic method based on linguistic markers to assess whether specific hypotheses discussed in scientific writings are considered by the authors to be more or less reliable.

(*Thesis1*) More specifically, the use of complex arguments in scientific research articles argues against the possibility of using the *quantity* of certainty/uncertainty markers – i.e. statistical analyses on the frequency

⁶ See Harman 1965; Boyd 1991; Psillos 1999; Ladyman 2002.

of their use – to draw conclusions on whether specific articles are more or less speculative or whether they adopt a more or less assertive/cautious style or whether the style of scientific articles has changed over time becoming more or less assertive/cautious (see e.g. Bongelli et al. 2014a; Bongelli et al. 2014b; Rosenthal et al. 2010; Kiyavitskaya et al. 2005; Vincze et al. 2008; Szarvas et al. 2008). Indeed, this last conclusion is coherent with the findings of e.g. Pinto et al. 2014 who do not notice any significant variation in the frequency of the use of certainty/uncertainty markers in medical papers of different periods, and conclude that – even though there are differences among disciplines in their use of hedges/boosters – in general the adoption of a more cautious or confident style by the authors depends on their individual preferences.

These considerations on the discrepancy between the certainty/uncertainty of the various claims made in an argument and its strength offer a *theoretical* argument against the possibility of using the frequency of automatically detected certainty/uncertainty markers to draw conclusions on the text, and on whether its style is cautious rather than assertive. However, it is still possible to argue for a different hypothesis: one could consider individual parts (claims) of a text – be they premises or conclusions of the arguments that compose it – and maintain that, if they are hedged, they must be considered speculative, while if they are boosted (or at least do not include hedges), they must be considered factual.

At first glance, the idea that an author hedges a claim s/he makes when s/he wants to express his/her uncertainty about it and that, on the contrary, an author boosts (or at least does not hedge) a claim s/he makes when s/he wants to express his/her certainty about it seems to be obvious, but in fact, the matter is not so straightforward. As we discussed in the previous section (point b), literature on hedging and boosting devices points out that authors use them for a range of rhetorical reasons other than the specification of their degree of commitment to the truth of the positions they discuss. Argument analysis can help to clarify this point because it offers an independent means to assess whether an author is more or less certain about the various claims s/he makes and thus it can help in determining whether in general hedges and boosters are used consistently and reliably in arguments to indicate whether a premise or a conclusion is considered to be certain or uncertain.

We tried to address this issue empirically analyzing a sample of research articles and comparing the arguments devised by the authors from the point of view of their strength on the one hand and the hedging and boosting strategies they adopt on the other. To have a variegated pattern of argumentation strategies, we consider articles in the field of psychiatry, which is a particularly interesting discipline for this kind of discussion for a number of reasons. On the one hand, psychiatry is linked to medicine, belongs to the biomedical sciences and relies, at least in part, on the same empirical methods and techniques; however, on the other hand, it is also strongly linked to psychology and to philosophy and it makes extensive use of argumentation to assert its theses. Furthermore, its theories and hypotheses have often been subject to strong disagreements among scholars and to radical revisions over time. Because of these characteristics, psychiatric research articles – especially when they consider controversial topics – are particularly interesting with respect to the issue of certainty/uncertainty in scientific communication.

To make sure that the articles are homogeneous and thus synchronically comparable with respect to their definition of the psychiatric approach, and that the differences among them can be traced back solely to factors related to personal divergences on the suggested hypotheses, we chose only papers published in a single Journal: *The British Journal of Psychiatry* (BJP). BJP was originally founded in 1853 and has continued to publish without interruption up until the present⁷. This is considered as one of the world's leading psychiatric journals and it represents a link between European and North-American research. Since we are especially interested in complex arguments which also contain refutations and confirmation through refutation, we decided to consider a subclass of psychiatry articles focusing on a particularly controversial topic which underwent many radical revisions and changes over time, and where any revision or change presupposed the refutation of previous points of view (for a general philosophical perspective on how scientific research proceeds during scientific revolutions see Thagard 1992).

⁷ From 1855 till 1858 the periodical was known as *Asylum Journal* and from 1858 to 1863 it took the name of *Journal of Mental Science*.

The topic we consider most suitable for this aim is homosexuality. Homosexuality was pathologized in the 19th century and thus included in psychiatric theory and in the psychiatric lexicon as a form of mental ‘illness’ (Beyer 1987; Rosario 2002). Its official classification by the psychiatric community has changed radically over time until 1987 when it definitively ceased to be considered as a psychiatric pathology (Drescher and Merlino 2007; Porcher 2014: 27-28). Its psychiatric description along with its changes across time were codified in the five *Diagnostic and Statistical Manuals of Mental Disorders* (DSMs) published by the American Psychiatric Association from 1952 till 2013. These define the standard criteria for the classification of the mental disorders accepted at the time of publication of the respective manual⁸. In the first DSM edition (DSM-I) published in 1952 homosexuality was described as a sociopathic personality disturbance. In the second DSM edition (DSM-II) whose first edition appeared in 1968 sexual ‘deviations’ were no longer considered as a sociopathic personality disturbance, but were categorized as sexual ‘deviations’ due to a personality disorder. Only in the seventh printing of DSM-II in 1974 homosexuality ceased to be interpreted as an ‘illness’ and started to be considered as a ‘sexual orientation disturbance’: in virtue of this definition homosexuality was considered as a sexual disorder only in those cases where people experienced a conflict with their sexual orientation. In the third DSM edition which appeared in 1980 (DSM-III) this classification was adjusted again and the label ‘sexual orientation disturbance’ was changed to ‘ego-dystonic homosexuality’ to indicate that sexual orientation might become a disorder in cases where it is not compliant with one’s ideal self-image. Since the issue of one’s ideal self-image appeared to be due to cultural factors while the reason why one might experience his/her homosexuality as ego-dystonic could be traced back to the widespread cultural homophobia, this classification was canceled in the revision of DSM-III-R published in 1987. DSM-III-R only mentions “sexual disorders not otherwise specified”, without considering homosexuality as a particular category. As specified by DSM-IV (1994) the diagnosis of “not otherwise specified disorders” had to rely on symptoms that “cause(s) clinically significant distress or impairment in social, occupational, or other important areas of functioning” (DSM-IV 1994: 7). Since DSM-IV homosexuality has definitively been depathologized and in the revision of 2000 “reparative”, “conversion”, and/or “aversion” therapies were condemned.

115 articles considering (more or less specifically) homosexuality (till the beginning of the 20th century also called “inversion”) have been published in JBP since its foundation in 1853 up until today. 26 of them are research articles; the others are brief articles (mainly, medical notes or literature reviews). We focused on the research articles which were published in 1921, 1957, 1962, 1964 (one article for each year), 1965 (two articles), 1968 (one article), 1969 (3 articles), 1970, 1971, 1972 (one article for each year), 1973 (3 articles), 1974 (3 articles), 1980 (2 articles), 1981, 1983, 1986, 1999, 2001 (one article for each year).⁹ We carried out

⁸ DSM-I was published in 1952; DSM-II appeared in 1968, but it underwent some important changes in the seventh printing of 1974; DSM-III was published in 1980 but in 1987 a revisited version (DSM-III-R) came out; DSM-IV appeared in 1994; a revisited version was then published in 2000. The fifth and last version of DSM was published in 2013.

⁹ Here we report the complete list of the titles and authors making up the sample. 1921 (*Homosexuality*, by C. S. Read); 1957 (*Psychometric Aspects of Homosexuality*, by T.G. Grygier); 1962 (*Homosexuality and Genetic Sex* by M. Pritchard); 1964 (*Homosexuality in Twins: A Report on Three Discordant Pairs* by N. Parker); 1965 (2 articles of the same author, *On the Genesis of Female Homosexuality* and *On the Genesis of Male Homosexuality: An Attempt at Clarifying the Role of the Parents* by E. Bene); 1968 (*Studies in Female Homosexuality IV. Social and Psychiatric Aspects* by F. E. Kenyon); 1969 (*Parental Age of Homosexuals* by K. Abe and P. A. P. Moran); 1969 (*Aversion Therapy of Homosexuality. A pilot study of 10 cases* by J. Bancroft); 1969 (*Homosexuality, Exhibitionism and Fetishism-Transvestism. Some Experiences in the Use of Aversion Therapy in Male* by B. H. Fookes); 1970 (*Subjective and Penile Plethysmograph Responses to Aversion Therapy for Homosexuality: A Follow-up Study* by N. McConaghy); 1971 (*A Male Monozygotic Twinship Discordant for Homosexuality. A Repertory Grid Study* by K. Davidson, H. Brierley and C. Smith); 1972 (*Parent-Child Relationships and Homosexuality* by G. Robertson); 1973 (*Heterosexual Aversion in Homosexual Males*) by K. Freund, R. Langevin, S. Cibiri and Y. Zajac; 1973 (*Classical, Avoidance and Backward Conditioning Treatments of Homosexuality* by N. McConaghy and R. F. Barr); 1973 (*Doctors' Attitudes to Homosexuality* by P. A. Morris); 1974 (*Sex Chromosome Abnormalities, Homosexuality and Psychological Treatment* by A. Orwin, S. R. N. James and R. K. Turner); 1974 (*Parental Background of Homosexual and Heterosexual Women* by M. Siegelman); 1974 (*Personality Characteristics of Male Homosexuals Referred for Aversion Therapy: A Comparative Study* by R. K. Turner, H. Pielmaier, S. James and A. Orwin); 1980 (*Social and Psychological Functioning of the Ageing Male Homosexual* by K. C. Bennett and N. L. Thompson); 1980 (*Homosexuality and Parental Guilt* by B. Zuger); 1981

(manually) two separate analyses on this data. (a) For the first analysis, we identified the linguistic lexical or morphosyntactical markers that are commonly used to assess the certainty/uncertainty of texts and speeches (Hyland 1996b; Hyland 2004: 188-189). (b) For the second analysis, we used basic tools from informal logic to reconstruct the structure of the argument in each article. Specifically, we identify the basic and the non-basic premises as well as the intermediate and the final conclusions, rewriting the extended argument presented in each paper in standard form and diagramming it in order to make its structure clearly visible. These two methods of analysis were then compared in order to show that they produce diverging results.

(Thesis2) The result we obtained from this comparison was fairly homogeneous for all the articles we considered and it univocally suggests that the use of hedging and boosting devices is not consistent with respect to the degree of certainty that emerges from the overall argument presented by the papers and that hedges and boosters are not used only or primarily to indicate certainty/uncertainty but fulfill different rhetorical aims. Thus, it would be a mistake to trust the linguistic markers of certainty and uncertainty to distinguish factual from speculative knowledge.

For reasons of space it would be impossible to illustrate this comparison for all the articles we considered. However, it is important to give a precise idea of how hedging and boosting strategies are used in research articles as well as to specify how argument analysis can be concretely applied with the aim of determining how certain/uncertain (i.e. strong/weak) arguments presented by authors in scientific articles may be. Thus, in the following section of this work we will first focus on three of the articles we selected which were published in critical years for psychiatric research on homosexuality (1957, 1962, 1974).

We should point out that this selection of articles is not meant to be an actual 'corpus' since the set of articles we analyzed (26 research papers) is much too small for this aim. This selection is intended to give some significant or prototypical examples of research writings in which the degree of certainty/uncertainty signaled by lexical markers do not correspond to the degree of certainty with which an author asserted a position as assessed on the basis of an independent evaluation of the strength of the arguments and sub-arguments developed in the article. In this sense, the selected papers should be considered more as a list of meaningful counterexamples which challenge (falsify) the claim that linguistic markers can be used as an automatic means to determine the degree of certainty attributed by authors to a certain thesis in corpora of *scientific research papers*.

4. Certainty identification – arguments vs. un/certainty markers: some psychiatry research papers

The first article we analyzed, written by T.G. Grygier, is dated 1957, and is on the *Psychometric Aspects of Homosexuality*. Its aim is "to discuss some attempts at measuring the psychological characteristics of the homosexual, and the direction and intensity of his impulses." (Grygier 1957: 514). The article addresses several psychological tests that "have been used in the diagnosis of homosexuality and in research concerned with this problem" (Grygier 1957: 514) and must be considered as extremely revolutionary for its time because the underlying position that motivates it is that homosexuality is not an illness for which we can find an etiology, but it is a personality trait and it has to do with personal history and personal relationships (Grygier 1957: 522). This is the reason why psychometric tests alone are considered an inadequate means for its assessment.

Here we address in particular two parts of this article. (A) First of all, we analyze the first argument presented by the author which concerns one of the most frequently used tests in clinical practice, i.e. the Minnesota Multiphasic Personality Inventory. (B) Secondly, we consider the conclusions of the paper, in

(*Neuroendocrine Mechanisms and the Aetiology of Male and Female Homosexuality* by M. J. MacCulloch and J. L. Waddington; 1983 (*Homosexuality and Lesbianism* by D.J. West); 1986 (*Homosexuality in Monozygotic Twins Reared Apart* by E. D. Eckert, T. J. Bouchard, J. Bohlen and L. Heston); 1999 (*British psychiatry and homosexuality* by M. King and A. Bartlett); 2001 (*Straight talking: an investigation of the attitudes and practice of psychoanalysts and psychotherapists in relation to gays and lesbians* by A. Bartlett, M. King and P. Phillips).

which the author summarizes his criticisms of the various tests he considers and the conclusions that can be drawn from them with respect to psychometric measures of homosexuality.

(A) The first argument reads as follows (hedges are in bold, there are no boosters):

“Among the inventories now used in clinical practice the most popular is **most probably** the Minnesota Multiphasic Personality Inventory [...]. There is a vast literature on its uses and abuses, and conflicting results about its factorial composition and validity have been reported [...]. The masculinity-femininity scale of the M.M.P.I. gives a particular picture of **supposedly** masculine characteristics.

A ‘typical’ male comes out as a cynical opportunist, prejudiced and superstitious, spiteful and revengeful, tough and cruel to animals, a slow thinker but quick to start a fight, enjoying primitive pleasures and having no manners.

By contrast, an effeminate man (or a woman) is sentimental, has aesthetic, artistic and cultural interests, is concerned with feelings and has insight into them, but also suffers from phobias, anxieties and worries.

In spite of the empirical basis of the test one **cannot escape a suspicion** that the above pictures represent stereotypes rather than reality. They are, moreover, Mid-Western American stereotypes which follow the familiar pattern of cultural superiority of the female sex. It **seems** that the masculinity-femininity scale has been so loaded with cultural factors that people with high intelligence and good education are bound to appear rather effeminate in the test profiles” (Grygier 1957: 515-516).¹⁰

This argument is a mode of refutation in which the author shows that the scale he addresses has absurd implications (because it depicts all people with high intelligence and good education as effeminate) and is therefore not reliable for clinical practice. The premises provide very strong evidence for the conclusion and the conclusion comes out as very certain. However, from the point of view of the style and especially the hedges used this argument it is nearly impossible to appreciate the strength and indisputability of its conclusion. Uncertainty markers only suggest that the masculine characteristics implied by the test *might be disputable*; that they *might be* due to prejudices; and that *probably* the masculinity-femininity scale is loaded with cultural factors. All the main points addressed are associated with uncertainty markers (we have seven claims, four uncertainty marker, no boosters). However, the analysis of the argument makes it clear that the author is very certain of all the points he makes and of his conclusion which is asserted without reserve or doubts. The reason why the author might have hedged claims that he considered to be perfectly certain can be traced back to *irony*. The absurd implications of the scale are ironically stressed using uncertainty markers.

(B) The analysis of this individual argument also provides us with a basis for addressing the overall argument presented in the paper which is summarized by the author in the concluding part and is quoted in full below (hedges are in bold, boosters are underlined; the numbers indicate the parts of the arguments: they are put in standard form and are diagrammed below).

“(i) It **appears** that no exact measures of the direction of sexual attraction are available at present [1]. Some indications of homosexual attraction or heterosexual revulsion **may** be obtained by means of projective tests, especially the Draw-a-Person Test, the Rorschach and the thematic tests [2], but the validity of the indicators is **very uncertain** [3].

(ii) With regard to the peculiarities of the homosexual's attitudes, behavior and personality traits more precise data are available [4]. It must be remembered, however, that the homosexual is not simply a man behaving like a woman [5], and for this reason tests of masculinity and femininity are **not always** very enlightening in the understanding of homosexual behavior [6]. As Terman and Miles (1936b) say, a ‘most emphatic warning is necessary against the assumption that an extremely feminine score for males or an extremely masculine score for female can serve as an adequate basis for the diagnosis of homosexuality, either overt or latent’. This is particularly true of high masculinity scores in women [7].

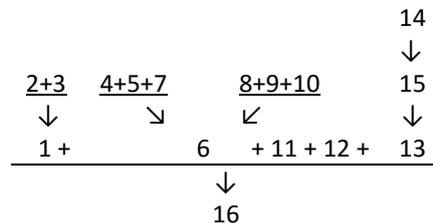
(iii) The concepts of masculinity and femininity are **not always clearly** understood [8]. Psychological masculinity is a more complex concept than the difference between the male and the female sexes [9]. The usual technique in measuring masculinity and femininity is a questionnaire, but more comprehensive tests, such as the Terman-Miles, or those measuring personality in developmental terms **may be** more promising [10].

(iv) All the projective and developmental tests **lay claim** to their ability to detect the dynamic aspects of homosexuality [11] and therefore to help in the understanding of the origin of sexual inversion in individual

¹⁰ The square brackets do not indicate omitted text, we left out only supporting literature; bold indicates the hedges.

cases and **possibly** in more general terms [12]. **In our present state of knowledge**, it is difficult to validate these claims [13], as understanding extends beyond measurement [14], and therefore beyond the usual sphere of experimentation and validation [15]. Precision of measurement is an important asset in this field, but the clinical psychologist must be more than a mere psychometrician [16].” (Grygier 1957: 523-524)¹¹

This argument can be diagrammed as follows:



Point (i) presents a linked argument in which the conclusion [1] is derived by the combination of two premises [2] and [3]. Both premises and the conclusion are hedged using corresponding markers, even though the conclusion [1] (i.e. no exact measures of the direction of sexual attraction is available at present) is quite strongly supported by the arguments given in the previous part of the paper. Points (ii) and (iii) offer a converging argument which, starting from the linked premises [4]-[5]-[7] and [8]-[9]-[10] – lead to the conclusion [6] (i.e. tests of masculinity and femininity are not enlightening in the understanding of homosexual behavior). This conclusion is hedged even though it is derived from a number of premises that are mostly presented as certain (six premises/three hedges/one booster – if we consider the arguments presented previously in the paper, the hedged conclusion appears to be considered as very certain by the author). Point (iv) is more complicated and it offers both a chain argument (leading from [14] to [15] and then to [13]) and two additional premises [11]-[12] which – in combination with the conclusion of the other lines of reasoning – allow us to infer the final conclusion [16]. The intermediate conclusion [13] is hedged as it is part of the premise [12]. Only the final conclusion is boosted, even though the marker is not referred directly to a theoretical claim, but concerns the role clinical psychologists should take considering all the conclusions reached in the previous part of the text.

This final part of the paper could be discussed starting from several points of view. We will examine the most important ones. In general, the claims made in the premises and in the intermediate conclusions of the argument are expressed in a cautious manner. However, this does not mean that the author considers his position to be speculative rather than factual. On the contrary, his argument is extremely strong and his conclusion is asserted with absolute certainty. Since the final conclusion is reached through converging and linked arguments, in line with the thesis we presented in §3 we could say that the conclusion is asserted with a high degree of certainty in spite of the fact that the individual premises are hedged because the conclusion is supported by all of them jointly. However, in this case this explanation seems not to be appropriate and the hedging devices used in the premises seem mainly to express the style of the author. As a matter of fact, he puts forward very strong arguments (analogous to the one illustrated before in A) to argue for the limitations of the various psychometric tests; thus, the sub-conclusion of the first arguments that are then used as a premise for this final conclusion are considered by the author to be very certain. In spite of this, he makes use of hedges to summarize them in the final part of the paper. The point here is that the use of hedging and boosting devices in the text seems to be incoherent and does not correspond with the degree of certainty that can be inferred from the author’s arguments. The factual knowledge expressed by the premises of this final conclusion is hedged. If these premises were uncertain, the final conclusion should be uncertain too. But the conclusion is asserted with certainty, confirming that the hedges in the premises *de facto* do not indicate uncertainty. The point that can be drawn from these observations is that markers are not reliable indicators of the degree of un/certainty attributed by the author to the various parts of his argument.

¹¹ In the original article the author uses Arabic numerals (1)-(4) instead of Roman ones (i)-(iv). We changed this in order to distinguish more clearly between the points made by the authors and the numbers used to diagram the argument.

If we consider the number of hedges and boosters used in the parts of the article that we could not report in full – i.e. the parts in which the preliminary conclusions have been reached – we notice that globally the balance is slightly in favor of hedges *H* over boosters *B* (2*H* vs. 7*B* for [1]; 7*H* vs. 3*B* for [2]; 8*H* vs. 2*B* for [3]; 9*H* vs. 7*B* for [4]) even though all the arguments that lead to these preliminary conclusions are quite strong. In fact, hedges are particularly high in the argument that leads to the conclusions [3] and [4] in which empirical data resulting from the application of the test are reported. This is very surprising because these data should univocally express factual knowledge and therefore be presented using certainty markers. On the contrary, they are presented in a hedged form as tendencies, while only the very final sentence which summarizes their meaning is expressed with a high degree of certainty: “these results do confirm the findings of nuclear sexing and afford further direct evidence against Lang’s hypothesis.” (Pritchard 1962: 622)

As for the final part of the article reported in full above, the premises of the argument – which are the sub-conclusions reached in the previous sections – constitute a *crescendo* of boosting strategies. The first premise is mildly hedged, but refers to the evidence which supports Lang’s theory indicating that it is weak. Premise [2] includes one booster in the first part of the claim, but the positive statement made in the second part using the indicative mood is also presented as factual. Premise [3] is boosted with two markers while premise [4] includes one booster only, even though it is presented univocally as factual through the specification that there is direct evidence for it. In sum, the premises give very strong support for the conclusion, which is asserted with a high degree of certainty.

The problem with the use of these hedging strategies is that they are used incoherently in the paper in at least two different ways. Factual knowledge (empirical data) is presented as uncertain. Moreover, the very same claims are presented as uncertain in the body text where they constitute the intermediate conclusion reached in the various sections, while they are presented as certain when they are used as premises to draw the final conclusion in the ‘Summary and conclusions’ section. A possible explanation of this inconsistent use of hedging strategies might be that the author wanted to be careful and cautious in criticizing Lang’s view which in 1962 was still a well-established and widely accepted position. Since the empirical data were the strongest and most factual element against this view, he is most careful when he is presenting them. Whatever the reasons which motivated the author, what we have in this article is an incoherent use of hedging strategies which would be completely misleading if we tried to rely on it to assess what knowledge the author considers to be factual or instead speculative.

The third article we would like to directly address here was written by Orwin and collaborators in 1974 and is on *Sex Chromosome Abnormalities, Homosexuality and Psychological Treatment*. The article discusses one single case of sexual reorientation of a male homosexual with Klinefelter’s syndrome using electric aversion therapy.¹³ Klinefelter’s syndrome is a chromosomal condition that leads to hypogonadism and other analogous dysfunctions. Most patients have less sexual interest; when they do, they are usually heterosexuals, confirming that homosexuality is not due to a chromosomal condition. The reason why the authors consider the case of their homosexual patients with Klinefelter’s syndrome to be particularly interesting is that the psychological treatment (i.e. aversion therapy) was successful in spite of the patient’s condition confirming that “abnormal sex chromosomes are relatively unimportant as determinants of psychosexual disorders.” (Orwin et al. 1974: 295)

The article cannot be reported in full. For this reason, we will focus on the final part of the *Discussion* in which the authors summarize the core of their argument (hedges are in bold, boosters are underlined; the numbers indicate the parts of the arguments: they are put in standard form and are diagrammed below):

¹³ While this was considered a pathology, a wide variety of techniques had been used in the treatment of homosexuality. Among them there were several types of “aversion therapies” which made use of various kinds of aversive stimuli to change the preference of so-called inverts towards heterosexuality. In this paper, the authors used an ‘anticipatory avoidance therapy’ analogous to that described by MacCulloch and Feldman (1967: 594). Basically, this is a technique similar to classical conditioning: the patient receives an electric shock when he is watching at pictures of attractive males.

“The picture from these descriptive papers is of considerable variation but of no **definite** correlation between Klinefelter's syndrome and specific psychosexual disorders [1]. [...]”¹⁴

In this case of Klinefelter's syndrome there was in fact ample evidence to **suggest** that the homosexuality was psychogenic in origin (e.g. early absence of father figure; maternal dominance causing repression of masculine assertiveness; seduction by stepfather; and small testes leading to feelings of sexual inadequacy with consequent avoidance of women) [2]. It **could** therefore be hoped that the sexual orientation would be altered by psychological treatment despite the presence of the sex chromosome abnormality, and there was a successful outcome [3]. This **tends** to confirm the relative unimportance of this abnormality, at least in the development of sexual preferences [4], and **may** indicate more generally that such abnormalities present no bar to the psychological treatment of psychosexual disorders [5].” (Orwin et al. 1974: 294)

The authors propose essentially a chain argument of the form:



Each step of the argument consists of views which are well-established in the literature and depend more or less directly on the refutation of Lang's argument which was already discussed on the basis of the 1962 article. As Orwin and colleagues acknowledge already in the very first sentence of their article: “The theory that some male homosexuals might have female chromosomal constitution (Lang 1940) became untenable when techniques for studying sex chromosomes were developed.” If this position is untenable, then it must be certain that sex chromosome abnormality and homosexuality are not related and that homosexuality has a psychogenic origin. In spite of the fact that at the time in which the authors write this appears to have been factual knowledge, their argument is studded with hedges. All steps of the argument present one uncertainty marker, including the last conclusion. This is very surprising since, in the light of the other positions discussed in the paper and of the references quoted in it, all the sub-conclusions of this argument as well as the final conclusion are considered as factual knowledge and should be asserted with a high degree of confidence.

The reason why they are hedged can probably be traced back to the authors' intent to make their hypothesis more appealing. Indeed, the point the authors want to make in their paper is quite predictable and unsurprising considering the knowledge available at that time; however, hedging strategies are used to artificially create some 'suspense'. Thus, the authors seem to use hedges as a rhetorical strategy to make normal science and its activities of puzzle-solving (in Kuhn's sense) appear more extraordinary. Also in this case, it would be a mistake to trust the linguistic markers of certainty and uncertainty to distinguish factual from speculative knowledge (*Thesis2*).

In general, our analysis on the whole sample shows that, when research challenges a well-established position, hedges are used primarily with the aim of being cautious, to avoid personal commitment or analogous (in our sample this seems to be the case for Bancroft 1969; Fookes 1969; McConaghy 1970; McConaghy and Barr 1973; Freund et al. 1973). On the contrary, when a paper discusses a view that can be considered as mainstream, hedges are mainly used in order to increase the interest of the reader and to make the results more appealing (in our sample this is the case for Davidson and Brierley 1971; Orwin et al. 1974).

Furthermore, there is another element that stands out in the analysis we carried out on our sample. Indeed, this element shows that the so-called paratext – e.g. tables, diagrams, figures etc. – is essential for understanding the authors' hedging and boosting strategies (Lemke 1998; Pinto et al., 2014). Our analysis reveals that the use of non-linguistic devices in papers became increasingly frequent in more recent times confirming that paratextual (non-linguistic) structures gained importance in psychiatric articles when data

¹⁴ In the text we left out the authors merely present further literature which confirms the point.

and empirical proof started to be considered as central also for this discipline, to the detriment of its connection with the human sciences. Initially used as a support for the text, with time they become part of the text and of its argumentative structure. In the sample we considered (which includes articles largely written during a period of time from the end of the 60s to the beginning of the 80s) the function of paratextual elements (above all tables) is to present the available data as evidence that speaks for itself. Unexpectedly, the increase in the use of these means is accompanied by an increase in the use of hedging strategies and especially of lexical uncertainty markers. In this way, the style becomes more impersonal, the authors do not take a binding position and their participation in what they report is neutralized. Thus, the hedging strategies became a 'distinctive mark' of scientific writings, in line with a Weberian view of science and of scientific communication, according to which scientific description should be as objective as possible and therefore should not depend on some preconception or theoretical conviction of the researchers (Crismore and Farnsworth 1990: 135).

In the articles we analyzed, the increased use of both hedging devices and tables is not aimed at indicating that the authors do not consider their results to be factual or certain (Bene 1965a; Bene 1965b; Kenyon 1968; Freung et al. 1973; McConaghy and Barr 1973; Siegelman 1974). It is rather meant to minimize their responsibility for what they report, making the scientific results appear to be independent of the views of the persons who report them. This use of hedging devices in research which relies on empirical results presented through the paratext shows once more that hedging (and boosting) devices cannot be considered as reliable indicators of the degree of certainty/uncertainty attributed by the authors to the hypotheses they consider. Indeed, this confirms that they perform a primarily rhetorical function. In this sense, an actual evaluation of the argument is the only means we have to determine whether the knowledge conveyed in a scientific paper is meant by the authors to be factual or speculative.

5. Concluding remarks

The issue we address in this work is whether the analysis of linguistic markers is a suitable method to assess the degree of certainty/uncertainty attributed by authors to the scientific hypotheses they discuss. We focus on scientific research articles only, which we consider to be both more difficult and more interesting to work with than other kinds of texts such as popular science texts or short research reports. The reason for this is that they do not merely illustrate specific scientific conclusions, but rather present complex arguments in which the final position is supported on the basis of many kinds of elements: positive evidence as well as refutations of conflicting, alternative views.

The hypothesis we started with in this paper is that the number (the frequency of use) and the type of the markers occurring in a paper (hedges vs. boosters) does not offer any reliable index to assess whether the arguments put forward by authors in their papers are more or less strong (i.e. whether the authors are more or less certain of the hypotheses they discuss in their scientific articles). The analysis we carried out on the sample of 26 articles on homosexuality taken from the *British Journal of Psychiatry* and exemplified by the three papers discussed in the previous section fully confirm this hypothesis. In fact, we showed that there is no relationship between the linguistic markers used in the various steps of an argument and its strength or the actual certainty of the premises and conclusions presented in it.

Hedges are considered to be more important than boosters in scientific literature because often to express factual knowledge in science we simply make claims in the indicative mood (see above §2). However, it is exactly these hedging kinds of markers which appear to be most problematic. In fact, our analysis confirms that an extensive use of hedges does not necessarily or even usually indicate uncertainty towards a position or a cautious commitment to its truth. Hedges are used for a number of other rhetorical reasons. Beyond the reasons that are most often pointed out in the literature (see above §2), the examples we discussed show that hedging strategies might also be used to make an argument more interesting (less routine and/or more extraordinary) or to be ironical and to show e.g. how absurd the implications of a certain viewpoint may be (on this aspect of scientific communication see also Knorr-Cetina 1981; Gilbert and Mulkey 1984; Latour and Woolgar 1979; Liddicoat 2005) or to make scientific results appear to be neutral, objective and independent of the opinions/preconceptions of the authors.

As for the project of developing an automatic method based on the search for hedges/boosters which could assess the commitment of authors to the truth of the theses they discuss, our analysis indicates that this is doomed to fail at least if the idea is to use such a method with scientific research articles, which are arguably the most interesting target for such an application. Since scientific research involves complex arguments devised to support specific conclusions, the automatic analysis of lexical and morphosyntactical markers is of no help in assessing the degree of certainty with which specific positions are affirmed, because there is no correspondence between the number and kinds of markers used and the strength of the argument. This has various implications for various relevant issues. First of all, linguistic markers alone are not an adequate way to distinguish factual and speculative information in a text. Secondly, an approach based on these markers is inadequate to track down the thinking processes underlying scientific reasoning. Finally, assessing the degree of certainty/uncertainty with which a hypothesis is put forward in a scientific paper can only be done on the basis of the (manual) reconstruction of the arguments it presents. Moreover, using this method it is also possible to establish the basis for this certainty/uncertainty and this is an essential piece of information in the dialectic and cooperative dynamic of science.

6. References

- Afsar, Hassan Soodmand, Mohamad Moradi and Raouf Hamzavi. 2014. Frequency and Type of Hedging Devices Used in the Research Articles of Humanities, Basic Sciences and Agriculture. *Procedia. Social and Behavioral Sciences*, 136: 70-74. Doi: doi:10.1016/j.sbspro.2014.05.290
- Agarwal, Shashank and Hong Yu. 2010. Detecting hedge cues and their scope in biomedical text with conditional random fields. *Journal of Biomedical Informatics*. 43: 953-961. doi: 10.1016/j.jbi.2010.08.003.
- Aguilar, Marta. 2008. *Metadiscourse in Academic Speech: A Relevance-Theoretic Approach*. Berlin: Peter Lang.
- Aikhenvald, Alexandra Y. 2003a. Evidentiality in typological perspective. In *Studies in Evidentiality*, ed. A.Y. Aikhenvald, 1-31. Amsterdam/Philadelphia: John Benjamins Publishing.
- Aikhenvald, Alexandra Y. 2003b. Preliminaries and key concepts. In: *Evidentiality*, ed. A.Y. Aikhenvald, 1-22. New York: Oxford University Press.
- Auger, A. and J. Roy. 2008. Expression of Uncertainty in Linguistic Data. *Information Fusion, IEEE*, 11th Conference on Information Fusion:1860-1866.
- Bednarek, Monika. 2006. Epistemological positioning and evidentiality in English news discourse: A text-driven approach. *Text & Talk*. 26(6):635-660.
- Beyer, Ronald. 1987. *Homosexuality and American Psychiatry: The Politics of Diagnosis*. Princeton (NJ): Princeton University Press.
- Boas, Franz. 1938. Language. In *General Anthropology*, ed. F. Boas, 124-145. Boston/New York: D. C. Heath and Company.
- Bongelli, Ramona, Carla Canestrari, Ilaria Riccioni, Andrej Zuczkowski, Cinzia Buldorini, Ricardo Pietrobon, Alberto Lavelli and Bernardo Magnini. 2012. A corpus of scientific biomedical texts spanning over 168 years annotated for uncertainty. In *LREC*, 12: 2009-2014.
- Bongelli, Ramona, Ilaria Riccioni, Carla Canestrari, Ricardo Pietrobon, Andrej Zuczkowski. 2014b. BioUncertainty: A historical corpus evaluating uncertainty language over a 167-year span of biomedical scientific articles. In *Communicating Certainty and Uncertainty in Medical, Supportive and Scientific Contexts*, ed. A. Zuczkowski, R. Bongelli, I. Riccioni, C. Canestrari, 309-339. Amsterdam: John Benjamins Publishing.
- Boyd, Robert. 1991. Explanation, Explanatory Power and Semolicity. In *The Philosophy of Science*, ed. R. Boyd, R. Gasper, D. Trout, 349-377. Cambridge (MA): The MIT Press.
- Brown, Penelope and Stephen C. Levinson. [1978](1987). *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press.
- Coates, Jennifer. 1983. *The Semantics of the Modal Auxiliaries*. London-Camberra: Croom Helm.

- Chafe, Wallace and Johanna Nichols. 1986. Introduction. In (vii – xi). In *Evidentiality: The Linguistic Coding of Epistemology*, ed. W. Chafe, J. Nichols, vii-xi. Praeger (NJ): Ablex.
- Chao, Feng and Guangwei Hu. 2014. Interactive Metadiscourse in Research Articles: A Comparative Study of Paradigmatic and Disciplinary Influences. *Journal of Pragmatics*, 66: 15-31.
- Choi, Eunsol, Chenhao Tan, Lillian Lee, Cristian Danescu-Miculescu-Mizil and Jennifer Spindel. 2012. Hedge Detection as a Lens on Framing in the GMO Debates: A Position Paper. In *ACL-2012 Workshop on Extrapositional Aspects of Meaning in Computational Linguistics*: 70-79.
- Coffin, Caroline Jane, Mary Jane Curry, Sharon Goodman, Ann Hewings, Theresa Lillis and Joan Swann. 2003. *Teaching Academic Writing: A Toolkit for Higher Education*. London: Routledge.
- Cornillie, Bert. 2009. Evidentiality and epistemic modality. On the close relationship between two different Categories. *Functions of Language*. 16(1):44-62. doi 10.1075/fo1.16.1.04cor
- Crismore, Avon and Rodney Farnsworth. 1990. Metadiscourse in popular and Professional Science Discourse. In *The Writing Scholar: Studies in Academic Discourse*, ed. W. Nash, 119-136. Newbury Park (CA): Stage.
- Crismore, Avon, Raija Markkanen and Margaret S. Steffensen. 1993. Metadiscourse in Persuasive Writing. *Written Communication*, 10(1): 39-71.
- Dendale, Patrick and Liliane Tasmowski. 2001. Introduction: Evidentiality and related notions. *Journal of Pragmatics*. 33: 339-348. doi:10.1016/S0378-2166(00)00005-9
- Drescher, Jack and Joseph P. Merlino (2007). Introduction. In *American Psychiatry and Homosexuality: An Oral History*, ed. J. Drescher, J.P. Merlino, 1-11. New York/London: Harrington Park Press.
- Durik, Amanda M., M. Anne Britt, Rebecca Reynolds and Jennifer Storey. 2008. The Effects of Hedges in Persuasive Arguments. A Nuanced Analysis of Language. *Journal of Language and Social Psychology*, 27(3): 213-234.
- Farkas, Richard, Veronika Vincze, György Móra, Janos Csirik and György Szarvas. 2010. The CoNLL-2010 Shared Task: Learning to Detect Hedges and their Scope in Natural Language Text. In: *Proceedings of the 14th Conference on Computational Natural Language Learning: Shared Task*, 1-12.
- Gilbert, G. Nigel and Michael Mulkay. 1984. *Opening Pandora's Box: A Sociological Analysis of Scientific Discourse*. Cambridge: Cambridge University Press.
- Grygier, T.G. (1957). Psychometric Aspects of Homosexuality. *The British Journal of Psychiatry*, 103: 514-526. DOI: 10.1192/bjp.103.432.514
- Harman, Gilbert H. 1965. Inference to the Best Explanation. *Philosophical Review*. 74(1): 88-95.
- Hempel, Carl Gustav. 1965. *Aspects of Scientific Explanation*, New York: Free Press.
- Hempel, Carl Gustav and Paul Hoppenheim. 1948. *Studies in the Logic of Explanation*. *Philosophy of Science*, 15(2): 135-175.
- Heritage, John. 2011. Territories of knowledge, territories of experience: Emphatic moments in interaction, in: *The morality of knowledge in conversation*, ed. T. Stivers, L. Mondada, J. Steensig, 159-183. Cambridge: Cambridge University Press.
- Holmes, Janet. (1983). Speaking English with the appropriate degree of conviction. In *Learning and teaching languages for communication: Applied linguistics perspectives*, ed. C. Brumfit, 100-121. London: BAAL.
- Holmes, Janet. 1990. Hedges and boosters in women's and men's speech. *Language and communication*. 10(3): 185-205.
- Holmes, Janet. 1984. Modifying illocutionary force. *Journal of Pragmatics*. 8: 345-365.
- Hyland, Ken. 1995. Author in the text. Hedging scientific strategies. *Hong Kong Papers in Linguistic and Language Teaching*, 18: 33-42.
- Hyland, Ken. 1996a. Writing Without Conviction? Hedging in Scientific Research Articles. *Applied linguistics*, 17(4): 433-454.

- Hyland, Ken. 1996b. Talking to the academy: Forms of Hedging in Science Research Articles. *Written Communication*. 13(2): 251-281.
- Hyland, Ken. 1997. Scientific Claims and Community Values: Articulating an academic culture. *Language and Communication*. 17(1): 19-32.
- Hyland, Ken. 1999. Academic Attribution: Citation and the Construction of Disciplinary Knowledge. *Applied linguistics*. 20(3): 341-367.
- Hyland, Ken. 2002. Authority and invisibility: Authorial identity in academic writing. *Journal of Pragmatics*. 34: 1091-1112.
- Hyland, Ken. 2004. *Disciplinary Discourses: Social Interactions in Academic Writing*. USA: University of Michigan Press.
- Hyland, Ken. 2009. Writing in the Disciplines: Research Evidence for Specificity. *Taiwan International ESP Journal*. 1(1): 5-22.
- Hyland, Ken. 2011. Disciplines and Discourses: Social Interactions in the Constructions of Knowledge. In *Writing in the Knowledge Society*, ed. D. Starke-Meyerring, A. Paré, N. Artemeva, M. Horne, L. Yousoubova, 193-214. West Lafayette (IN): Parlor Press.
- Hyland, Ken. 2012. *Disciplinary identities: individuality and community in academic discourse*. Cambridge, (UK): Cambridge University Press.
- Kamio, Akio. (1997). *Territory of information*. Amsterdam: John Benjamin.
- Kim, Jin-Dong, Tomoko Ohta, Sampo Pyysalo, Yutaka Kano and Junichi Tsujii. 2009. Overview of BioNLP'09 Shared Task on Event Extraction. *Proceedings of the BioNLP 2009 Workshop*. Boulder, Colorado, 1-9.
- Kiyavitskaya, Nadzeya, Nicola Zeni, James R. Cordy, Luisa Mich and John Mylopoulos. 2005. Semi-automatic Semantic Annotations for Web Documents. *Proceedings of SWAP 2005*, 14-15.
- Knorr-Cetina, Karin. 1981. *The Manufacture of Knowledge: An Essay on the Constructivist and Contextual Nature of Science*. Pergamon: Oxford.
- Kuhn, T. 1970. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.
- Ifantidou, Elly. (2005). The Semantics and Pragmatics of Metadiscourse. *Journal of Pragmatics*. 37: 1325-1353.
- Ladyman, James. 2002. *Understanding Philosophy of Science*. London: Routledge.
- Lakoff, George. 1973. Hedges: A Study In Meaning Criteria and the Logic of Fuzzy Concepts. *Journal of Philosophical Logic*, 2: 458-508.
- Lakoff, George. 1987. *Women, Fire, and dangerous Things. What Categories Reveal about the Mind*. Chicago: Chicago University Press.
- Latour, Bruno and Steve Woolgar. 1979. *Laboratory Life: The Social Construction of Scientific Facts*. Beverly Hills (CA): Sage.
- Lemke, Jay. 1998. Multiplying Meaning. Visual and Verbal Semiotics in Scientific Text. In *Reading Science. Critical and Functional Perspectives on Discourse of Science*, ed. J.R. Martin, R. Veel, 87-113. London: Routledge.
- Liddicoat, Anthony J. 2005. Writing about Knowing in Science: Aspects of Hedging in French Scientific Writing. *LPS & Professional Communication*. 5(2): 8-27.
- MacCulloch, M.J. and M.P. Feldman. 1967. Aversion Therapy in Management of 43 Homosexuals. *British Medical Journal*, 2: 594-597.
- Morante, Roser and Walter Daelemans. 2009. Learning the scope of hedge cues in biomedical texts. *Proceedings of the Workshop on BioNLP*, 28-36.
- Myers, Greg. 1989. The pragmatics of politeness in scientific articles. *Applied Linguistics*, 10: 1-35.

- Origgi, Gloria and Dan Sperber. 2002. Evolution, Communication, and the Proper Function of Language. In *Evolution and the Human Mind: Language, Modularity, and Social Cognition*, ed. P. Carruthers, A. Chamberlain, 140-169. Cambridge (MA): Cambridge University Press.
- Orwin, Arnold, Sheelah R.N. James and Keith Turner (1974). Sex Chromosome Abnormalities, Homosexuality and Psychological Treatment. *The British Journal of Psychiatry*. 124: 293-295. DOI: 10.1192/bjp.124.3.293
- Pinto, Maria da Graça, Paulo Osorio and Fernanda Martins. 2014. A theoretical contribution to tackling certainty and uncertainty in scientific writing. Four research articles from the journal *Brain in focus*. In *Communicating certainty and uncertainty in medical, supportive and scientific contexts*, ed. A. Zuczkowski, R. Bongelli, I. Riccioni, C. Canestrari, 291-305. Amsterdam: John Benjamin.
- Porcher, Jose Eduardo. 2014. A Note on the Dynamics of Psychiatric Classification. *Minerva*. 18: 27-47.
- Pritchard, M. (1962). Homosexuality and Genetic Sex. *The British Journal of Psychiatry*, 108: 616-623. DOI: 10.1192/bjp.108.456.616
- Psillos, Stathis. 1999. *Scientific Realism*. Routledge: London New York.
- Rubin, Victoria L., Elizabeth D. Liddy and Noriko Kando. 2006. Certainty Identification in Texts: Categorization Model and Manual Tagging Results. In *Computing Attitude and Affect in Text: Theory and Applications*, ed. J.G. Shanahan, Y. Qu, J. Wiebe, 61-76. Dordrecht (NL): Springer.
- Rosario, Vernon A. 2002. *Homosexuality and Science. A Guide to debate*. Santa Barbara (CA): ABC-CLIO.
- Rosenthal, Sara, William J. Lipovsky, Kathleen McKeown, Kapil Thadani and Jacob Andreas. 2010. Towards Semi-automated Annotation for Prepositional Phrase Attachment. *Proceedings of the 7th Conference on International Language Resources and Evaluation LREC 2010*. <http://www.lrec.conf.org>
- Sinott-Armstrong, Walter and Robert J. Fogelin. 2009. *Understanding Arguments: An Introduction to Informal Logic*, Wadsworth: Cengage.
- Skelton, John. 1997. The Representation of Truth in Academic Medical Writing. *Applied Linguistics*. 18(2): 121-140.
- Szarvas, György, Veronika Vincze, Richard Farkas and Janos Csirik. 2008. The BioScope Corpus: Annotation for Negation, Uncertainty and Their Scope in Biomedical Texts. *Association for Computational Linguistics, BioNLP: Current Trends in Biomedical Natural Language Processing*, 38-45.
- Thagard, Paul. 1992. *Conceptual Revolutions*. Princeton: Princeton University Press.
- Thompson, G. (2001). Interaction in academic writing: Learning to argue with the reader. *Applied Linguistics*. 22: 58-78.
- Vincze, Veronika, György Szarvas, Richard Farkas, György Móra and Janos Csirik. 2008. The Bio-Scope Corpus: Biomedical Texts Annotated for Uncertainty, Negation and their Scopes. *BMC Bioinformatics*. 9(11): S9. DOI: 10.1186/1471-2105-9-S11-S9.
- Wilson, Deirdre. 2005. New Directions for Research on Pragmatics and Modularity. *Lingua*. 115: 1129-1146.
- Wilson, Deirdre. 2000. Metarepresentation in Linguistic Communication. In *Metarepresentations: A Multidisciplinary Perspective*, ed. D. Sperber, 411-448. Oxford: Oxford University Press.
- Winter, Stephan, Nicole C. Krämer, Leonie Rösner and German Neubaum. 2015. Don't Keep It (too) Simple – How Textual Representations of Scientific Uncertainty Affect Layperson' Attitude. *Journal of Language and Social Psychology*. 34(3): 251-272.
- Zuczkowski, Andrej, Ramona Bongelli, Ilaria Riccioni and Carla Canestrari. 2014. *Communicating certainty and uncertainty in medical, supportive and scientific contexts*. Amsterdam: John Benjamin.