

It Don't Mean a Thing... Simultaneous Interpretation Quality and User Satisfaction*

PHILIP MACDONALD

École des Hautes Études en Sciences Sociales, Paris

Abstract

The issue of quality has been extensively discussed in Interpreting Studies (IS). Quality is subjective, ineffable and cultural. As the “aspiring-to-science community” (hereafter “ATSC”¹) defines “scientific” as empirical, quantifiable and objective², it is bound to struggle when dealing with such a concept. Yet, precisely because it stipulates that a scientific approach requires a quantifiable dimension, it has to try and define quality in an objective manner. Shackled by its postulates, the ATSC has drawn upon two approaches that have predictably come short. One vainly seeks to define quality and subsequently “objective and quantifiable” criteria to assess it. The other claims to draw on marketing and strives to measure user satisfaction, primarily through questionnaires. The most advanced work in marketing, however, has taken on board the findings of cognitive

* I would like to dedicate this paper to one of my dearest friends, Alain Bonzon, who passed away not so long ago and happened to be an intensive Simultaneous Interpretation user. May he rest in peace.

1 I will use Gile’s convoluted description in this paper although I would prefer the word “positivists”.

2 A similar statement is unsurprisingly inscribed on the front of the Social Science Research Building at the University of Chicago. Frank Knight, the excellent albeit underestimated economist, is said to have remarked on it one day: “Yes, and when you can express it in numbers your knowledge is of a meagre and unsatisfactory kind!” (McClosky 1983: 482).

psychology and behavioural economics and rejects questionnaires as a reliable source of information about customers' thoughts and emotions. Moreover, IS questionnaires usually postulate clear-cut distinctions between such norms as fidelity to meaning and performance-related criteria. Quality is frequently restricted to performance in the booth although this could lead to significant problems being overlooked. Lastly, the ATSC might be stifling promising research approaches and projects because of its excessively restrictive criteria.

Introduction

Daniel Gile (1990a: 233) has argued that “a very important question which has never been studied in depth is the nature of interpretation quality” and called for research to look at how interpreters and delegates define quality. Eleven years later, he was still writing that “[a]ssessing quality in interpretation, as well as translation, evidently is a major issue, both at a professional as well as an educational level” (Gile 2001: 379). In fact, any discussion of expertise and professional performance requires shedding light on the concept of quality. Can “objective” criteria be defined? Or is there an intangible aspect that is subjective, because the rapport between interpreters and audience plays a crucial role? We have all experienced cases where users have praised our performance even though we know all too well that it was far from flawless and sometimes actually accumulated mistakes. In other words, what did the users like? If we know that we made numerous mistakes when interpreting names and figures, clearly we cannot hope to define quality by assessing our performance according to such criteria. I would argue that we are moving into a far more slippery area, where the users' reaction may very well depend on “subjective” factors such as the confidence conveyed by the interpreter's voice as well as the impression projected by the interpreter of being interested in the subject, and so forth.

The ATSC's determination to assess quality in interpretation and translation has to be understood in the light of their definition of “scientific”. Just as we need to define as precisely as possible a kilogramme, a metre or a litre in order to be able to make measurements, the rationale here is that, since we need to be able to quantify data when carrying out research, we must rely on an “objective and quantifiable” criterion, in other words quality. Afterwards, supposedly, we will be able to conduct tests to gauge the efficiency of, say, an approach in teaching as reflected by an improvement or deterioration in quality. Similarly, Gile (1990c: 29) has argued that scientific work is based on “facts collected through systematic observation, carefully checked and assessed”, to be contrasted with unscientific research based on “facts encountered in daily, personal and subjective experience”. His dichotomy is founded on the simplistic idea that facts exist objectively and simply wait for researchers to identify them. Such a pre-Kuhnian viewpoint stems from the inclusion of “empirical” in the definition of “scientific”. Here, we could draw on E.H. Carr's (1961: 23) famous metaphor when discussing such a viewpoint in history:

[...] facts are really not at all like fish on the fishmonger's slab. They are like fish swimming about in a vast and sometimes inaccessible ocean; and what the historian catches will depend, partly on chance, but mainly on what part of the ocean he chooses to fish in and what tackle he chooses to use – these two factors being, of course, determined by the kind of fish he wants to catch. By and large, the historian will get the kind of facts he wants.

I apologise for insisting on the fact that the ATSC definition of “scientific” is so restrictive but, as we shall see, this plays a crucial role in the debate and accounts for otherwise contradictory and inconsistent statements. An additional reason why quality has been widely discussed in IS is that we conference interpreters claim we deserve to be paid similar rates to those of highly qualified professionals and therefore need to justify such a demand. Any such research in IS therefore needs to be above suspicion like Caesar's wife in order to pre-empt accusations of manipulating findings³.

The foregoing leads to a paradox. Some twenty years ago, Gile hailed the fact that what he calls the aspiring-to-science community “has taken the lion's share both in recent publications and in conference participation” (Gile 1995a: 15). The reasons for its success were evident in his opinion: “In concrete terms, [ATSC scholars]:

- systematically conduct empirical testing of their ideas and theories;
- systematically provide evidence to back up claims;
- are explicit about their materials, methods and factual and/or logical grounds for their claims;
- make a clear distinction between documented facts and speculative thoughts” (Gile, 2004a).

Notwithstanding, although the changeover to a “scientific” approach could have been expected to result in major advances in the study of quality, the lack of any significant progress is striking. Kahane (2000) concedes that there is no consensus on “the elusive concept of quality; quality for whom, assessed in what manner?”. Collados Aís and her team (2007: 224) conclude their book by quoting Cartellieri's words from 1983: “Much still remains to be done to overcome the present unsatisfactory state of affairs in the sphere of reliable quality parameters”. In fact, such concepts as Quality or Beauty are ineffable, cultural and subjective. Seeking to define Quality “objectively” hardly makes more sense than trying to paint the wind. ASTC researchers have attempted to muddy the waters by talking about quality *per se* when they are clearly referring to a definition that necessarily includes a subjective dimension or about quality defined as equivalent to user satisfaction. In other words, the ASTC approach dooms research to a dead-end because of the flaws and shortcomings of its methodology.

3 I hasten to say that Gile is exemplary in terms of tackling head on the fact that research findings may be inconclusive or suggest that some of the claims made by interpreters are not justified. His willingness to discuss research that apparently undermines some long-standing beliefs and viewpoints in the interpreting community is highly commendable.

In my opinion, the fundamental fallacy of the ATSC is the conviction that “scientific” means objective, quantifiable and empirical. Anything in Interpreting Research that cannot be quantified is rejected into the purgatory of the “Liberal Arts Paradigm”. More fundamentally, although the ATSC portrays itself as scientific and objective, relying exclusively on solidly demonstrated evidence, its proponents repeatedly indulge in sweeping assertions that lack any basis whatever but seek to project the image of self-evident statements of fact. For instance, Gile (1995a: 20) writes:

Most interpreters and interpretation teachers with academic degrees in fields other than translation or interpretation are graduates of foreign languages and/or cultures (sic) departments. Only a few practi-searchers have a solid background in an *established scientific discipline such as linguistics or psychology* [...] (my italics)

On what grounds can Gile argue that linguistics and psychology are “established scientific disciplines”? Psychology was dominated for far too long by behaviourism, a school that incidentally also argued that its methodology was empirical, quantifiable and objective. Would Gile argue that sociolinguistics is scientific? Would he claim that Chomsky’s transformational generative grammar is undisputed and undisputable science? Likewise, Gile (2009: 144) writes: “scientific literature (...) *depends less on rhetoric*⁴ and more on *strict, systematic, cautious, logical, objective use of data* [than academic literature of the liberal arts type]” (my italics). To which it could be retorted that the ATSC resembles Molière’s Monsieur Jourdain who was astonished to learn that he had been speaking prose all his life: the ATSC apparently has been using rhetorical devices without being aware of it.

Take the habit of speaking about oneself or one’s work in the third person. McClosky (1990: 32-33) explains how “[t]he suppression of the “I” in scientific writing is more significant than one might think. [...] The scientist says: It is not I the scientist who makes these assertions but reality itself”. He adds that the underlying idea is that “[...] scientific texts are transparent, a matter of “mere communication”, [...], simply “writing up ‘theoretical results’ and ‘empirical findings’ (*ibid.* 36-37)”. Not all scientists feel the need to abide by such a convention with everything it entails. Indeed, Watson and Crick’s landmark paper begins: “We wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.)” (1953: 737).

By contrast, ATSC researchers have repeatedly used the third person to make unjustifiable value judgements as if they were perfectly bland and self-evident. For instance, Gile has argued that “most interpreters are humanities- or language faculty graduates with no scientific training and expertise, and *they find less motivation in the long, somewhat arid efforts actual research implies* in terms of data collection, analysis and tests of precise hypotheses than in free theorization” (my italics). Moser-Mercer (1994: 17) divides the IS community into followers of the “natural science paradigm”, where research requires “precision of logical processes” and a “liberal arts community” that is content with “explorations which *involve the intellect in a less logically rigorous manner*” (my italics). These two

4 Here, Gile is using a rather negative and limited definition of the word “rhetoric”.

statements hardly fit Gile's description of ATSC members' "reluctance to speculate [...] without the support of evidence for every step they take" (Gile 2004b): 126). Amazingly, he rejects Pöchhacker's description (2004: 109) of Moser-Mercer's statement above as "divisive" and disingenuously adds "until Pöchhacker presents evidence or at least a strong rationale to support this view, I should like to suggest the opposite"⁵. In other words, Gile seems to be saying that as Moser-Mercer is merely describing reality, i.e. the fact that the liberal arts community is content with explorations that are less taxing intellectually, who could object?

Likewise, Gile has also asserted that "questionnaires have been the most common means to determine user expectations and/or responses, as they are the most straightforward scientific way of collecting data on actual quality perception by delegates" (Gile 1991: 193-4). Kurz (2001: 397) has repeated this argument *verbatim* while referring to Gile's assertion. Neither writer provides the slightest shred of evidence to back up such a controversial claim (I would personally argue that a more compelling argument could be made in favour of participant observation) – although Gile defines the need to give references to "back statements which require them" as a "fundamental criterion for scientific quality" (1999a: 33). Another rhetorical device used at times consists in sprinkling a text with mathematical symbols even when this does not clarify the meaning of the text or provide any meaningful contribution to the argument. This list is obviously not exhaustive. The reader will find more examples of subjective (and even at times glaringly wrong) statements made by ATSC members masquerading as obvious statements of fact in the rest of this article.

1. Am I measuring quality or user satisfaction?

Some researchers have designed surveys targeting interpreters and not users. Such questionnaires seek to determine the importance of various components of quality in the opinion of interpreters. In particular, Chiaro and Nocella (2005: 177) contended that "a genuine delegate is likely to be hard put to be able to judge the fidelity of an interpreted speech with the original" and therefore sent their questionnaire to interpreters. Gile has repeatedly questioned the ability of users to assess interpretation quality criteria. In all likelihood, this is indeed frequently the case and delegates often lack any such expertise. Notwithstanding, the vast majority of users *do* assess, when filling in a questionnaire, the quality of interpretation.

As Kahane (2000) points out, Gile throughout his extensive body of work has been tireless in advocating the need for *objective, quantifiable* criteria. At the same time, Gile has repeatedly highlighted the problems encountered with respect to

5 Ironically, Gile (2004 *ibid*) then writes: "I should like to suggest that a fundamental difference between followers of the 'natural science paradigm' and followers of a 'humanities-inspired paradigm' (or whatever other names one might like to give them) is that the former tend to stay much closer to evidence in their inferences than the latter, for whom it is legitimate to interpret and extrapolate without necessarily having to justify every one of their assertions." Bear in mind that Moser-Mercer does not provide any evidence at all to back her somewhat contentious statement.

the subjective nature of assessment⁶. In fact, how could a consensus ever emerge on quality assessment guidelines? Grbić (2008: 234) states the obvious when she writes that “quality is not intrinsic to an object”. Luccarelli and Gree (2007: 2544) illustrate this point:

In 1996, the Private Market Sector invited an expert on quality control systems to a meeting in Lisbon to explore the possibility of AIIC or its members applying for ISO quality certification (...). Our quality control expert recognized the difficulty of the task. Quality would be difficult to control in conference interpreting precisely because it depends on so many factors beyond the interpreter. Listeners may very well form different impressions and give different evaluations. In other words, he deemed that a degree of subjectivity was unavoidable.

To give just one example, we sometimes adapt our output to the people listening to us when we know enough about them. I regularly interpret at a three-day management seminar held by a CAC 40 company. Twice, a former French rugby star was the guest speaker. The first time I was working for two Britons and an American who had all played rugby (the American amazingly had been a long-standing member of the US national team). The second time, I was interpreting for a German, a Dutchman and a Pole and safely assumed they knew virtually nothing, if anything, about the sport.

If I had been taped and the quality of my two versions had been assessed, the obvious and “objective” conclusion would have been that I made far more errors and omissions the second time around. If I had been taped only the second time, my rating would have been quite low. I can remember one specific example because a colleague asked me at the time how I had translated “*on ne va pas faire pleuvoir des chandelles*”. The first time I translated “we’re not going to kick up-and-unders all day” (trust me: that is one way of rendering what it means in English) and the second time “we’re not going to play negative rugby”. The first time I peppered my interpretation with rugby-specific jargon, the second time I made numerous omissions and actually “mistranslated” quite a few words while adding information the speaker had not given.

Likewise, Peter Mead (2005: 40) describes a study of how seven interpreting teachers at Italian and Austrian universities were asked to assess interpretations by five students:

Lack of consistency between the various assessments indicates considerable variability in standards and priorities from one assessor to another. It was emblematic, for example, that there was unanimity about awarding a pass or a fail for only three out of ten interpretations. Another interesting finding was that almost none of the seven assessors could generally be identified as a consistently higher (or lower) marker than others.

Any criticism on the grounds that the sample is too small statistically can easily be dismissed: extensive research in docimology has shown the subjective dimension of evaluation. This holds true in such subjects as mathematics as well

6 See Gile 1990b: 195 for instance. In particular, in this article Gile explores a wide range of reasons that rule out the idea of ever defining objective assessment criteria.

as physics. One examiner will fail a student who made a calculation error at the very outset and accordingly every other calculation was wrong, whereas another one will give the same student an excellent grade because she deems a silly mistake in an otherwise faultless demonstration relatively unimportant. Quality is by definition a subjective and multidimensional concept.

I will not discuss in depth Gile's "scientific" definition of interpretation quality as:

[...] a *subjectively* weighted sum of a number of components: the fidelity of the target-language speech, the quality of the interpreter's linguistic output, the quality of his or her voice, the prosodic characteristics of his or her delivery, the quality of his or her terminological usage, *all of them as perceived by the assessor*" (Gile 1995b: 151) (my italics).

He has even included a mathematical formula ($Q = \sum w_i c_i$) where c_i stands for the quality components and w_i for their relative weightings (ibid). He fails to explain, however, how these mathematical symbols enhance his explanation or how his approach can be reconciled with the inclusion of subjective components in his definition.

In sum, we are to build a rocket in order to explore a planet called Quality – although nobody has the slightest idea where the planet is to be found. Robert M. Pirsig in 1974 in his cult novel *Zen and the Art of Motorcycle Maintenance* summed up everything that can be said in this respect: even though Quality cannot be defined, you know what Quality is, in your opinion, when you see/hear/taste/read it⁷. The main character ends up on the verge of insanity because of his obsessive attempt to nail down the concept.

Another approach has been widely followed. Instead of pursuing the Holy Grail of defining quality, other researchers have sought to measure user satisfaction. Grbić (ibid: 236) has pointed out that several interpreting scholars explicitly refer to marketing and management studies, in particular Total Quality Management and Quality Assurance. An interesting point is that the main thinkers in these fields, the likes of Juran and Deming for instance, never define Quality – they simply measure the cost of failing to provide it. Citing Kotler and Armstrong⁸, Kurz (2001: 394) nonetheless asserts:

[...] quality must begin with customer needs and end with customer perception. There is no reason why this generally accepted marketing principle should not apply to conference interpreting as well

More accurately, it could be said that this principle is "generally accepted" among marketing apologists. Such an approach is blatantly functionalist. If quality is achieved by meeting or exceeding customer expectations, my children's subjective assessment of the quality of Big Mac hamburgers, i.e. outstanding, has to be

7 I would like to thank Tomás de la Guardia for reminding me of this book and making many other insightful suggestions.

8 Actually, this is a textbook for undergraduates.

endorsed. Hunger and thirst are physiological needs, but a craving for a Big Mac and a can of Coke is a want stimulated by massive advertising, which has resulted in a severe child obesity problem in many countries. How could Kotler's definition be of any use in terms of evaluating quality in interpreting?

The fundamental assumption accepted by nearly all IS researchers who believe quality reflects user satisfaction has been voiced, as said above, by Daniel Gile: "Questionnaires (...) are the most straightforward *scientific* way of collecting data on actual quality perception by delegates" (Gile 1991: 193-4) (my italics). Note that Chiaro and Nocella (2004) made a compelling case that most questionnaire-based quality research carried out in Interpreting Studies suffers from severe methodological flaws in terms of statistical methodology. Interestingly, in their original article they stated that "[w]e would like to approach the issue of quality from the angle of economics, bearing in mind that interpreting is a service [...]" (2004: 280). In their answer to Pöchlhammer, they (2005: 177) wrote "judging the quality of an interpretation is quite different from that of judging a regular marketable good. (We suggest that those convinced by our argument skip the rest of this section and move on to 3.2)". I beg to differ and strongly recommend not skipping the following part because it is illuminating:

A housewife asked to judge the quality of a pot of jam, for example, has a range of tangible and highly perceptible characteristics upon which to base her evaluation. The colour of the jam, how much it costs, it's (sic) shelf life, nutritional information on the label, packaging and, last but not least, it's (sic) flavour.

Several ideas spring to mind. First, a sleight of hand is obvious: we have moved from "the angle of economics", and a "scientific" approach buttressed by statistical concepts such as sum of the scores, non-comparative scales and chi squared testing to ... what could be best described as a naïve view of marketing. Second, the housewife (the authors' choice of word) they describe is a *femina oeconomica* who makes rational choices based on objective criteria and is fully aware of the reasons why she makes these choices.

Whatever economic woman may say when answering a survey, her choice is influenced by a myriad of factors she is unaware of. For instance, retailers use colours: red dominates at McDonald's because it causes you to eat more quickly, freeing up room for the next customer. Supermarkets are now using specialised scent machines wafting the smell of fresh baked bread and other scents because shoppers will spend more when smelling them. The best slots on shelves are at adult eye-level, and that is where relatively expensive products are put, often to the right of popular items (to increase the chances that right-handed shoppers will pick them up). The fresh fruit and vegetables section is systematically positioned next to the entry. As *The Economist* (18 December 2008: "The way the brain buys") points out:

For shoppers, this makes no sense. Fruit and vegetables can be easily damaged, so they should be bought at the end, not the beginning, of a shopping trip. But [...] selecting good wholesome fresh food is an uplifting way to start shopping, and it makes people feel less guilty about reaching for the stodgy stuff later on

In fact, the unreliability of surveys in terms of “reflecting” consumers’ feelings and, above all predicting subsequent actions, has been obvious for many marketing professionals for a long time. In the 1950s, for instance, Ernest Dichter⁹ and the Motivation Research school eschewed empirical marketing research and polling in favour of in-depth interviews and small panels. Robert Peterson (1992: 49) suggested in the 1990s that:

customer satisfaction ratings may well reflect the Hawthorne effect: Attempts to measure customer satisfaction will, in and of themselves, serendipitously increase satisfaction, regardless of the product or service being investigated.

Vavra (1997: 29) reported that “60% of all defecting customers were either extremely or very satisfied according to CSM data”¹⁰. American marketing specialist Jack Trout (2008: 42) pointed out that 89% of people who owned cars of a certain make said they were very satisfied and 67% said that they intended to purchase another car from that company. A follow-up study found that fewer than 20% actually did so. As Ogilvy Group UK vice chairman Rory Sutherland in Tarran (2011: *passim*) states:

No-one in any research group would ever say, “If there are four brands of shampoo, I’ll buy the one that has most bottles on the shelf”, or “I’ll choose the one that’s on the third shelf up because it’s the one that doesn’t require much reaching down” or “I’ll look at the prices of three products and choose the one in the middle.” In reality, we use heuristics and shortcuts and cognitively miserliness like this all the time.

Philip Graves (2010: 91-2) shows that the original estimate of consultants Deloitte & Touch projected that 12 million people would visit the Millennium Dome in London in 12 months. Advertising agency M&C Saatchi’s subsequent review argued, on the basis of surveys, that 12 million was a “conservative” figure. In fact, only 6.5 million visitors came – a disastrous shortfall. They failed to come despite what they had said when questioned about their intentions.

In fact, more and more companies and researchers agree that watching consumers behave is far more relevant than listening to what they say when asked questions. By the way, any teacher worth her grain of salt knows that asking pupils/students whether they have understood something is senseless. Not only will they feel reluctant to confess they do not understand something more than once from time to time, but also all too frequently they will sincerely and mistakenly answer in the affirmative. Graves (2010: 92-3) makes a telling point of obvious relevance for the issues discussed in this article:

Concern about the quality of research tends to be focused on the validity of the sample and the statistical significance of any differences in the data, but [...] the methodology can be pure and the results still grossly misleading.

9 He is credited with the concepts of focus groups and brand images.

10 CSM stands for Customer Satisfaction Management.

He carries on to describe in depth the flaws in surveys, focus groups and questionnaires used in marketing but his crucial point is as follows: “[.] over recent decades, a growing body of scientific evidence has revealed something that is both fascinating and somewhat disarming. We don’t think in the way we think we do” (*ibid*: x)¹¹. Sutherland spells out the logical conclusion that “survey research is an inherently unreliable means of getting to the truth of consumer behaviour and emotions.” For him (Tarran 2011) at the very most, “it’s still better than ignorance in many cases”.

2. Seventy times seven skins¹²

Cognitive psychology sheds light on the process with its “dual-process” model of the brain. According to Daniel Kahneman, who won the “Nobel Prize in Economics” in 2002, we use two fundamentally different modes of thought: “System 1¹³” and “System 2”. In his riveting book “*Thinking Fast and Slow*”, he writes that System 1 “operates automatically and quickly with little or no effort and no sense of voluntary control” while System 2 “allocates attention to the effortful mental activities that demand it, including complex computations” (Kahneman 2012: 20). He argues that System 1 is sensitive to subtle environmental cues, e.g. instantaneously detecting hostility in a voice, and enables us to carry out everyday activities, such as driving a car, without needing to focus on them and so forth. However, this system is terrible at the sort of statistical thinking required to make complex choices, jumps to wild conclusions and uses irrelevant mental shortcuts or heuristics¹⁴ and is prey to all sorts of illogical biases. System 2 is the conscious being we call “I” and mistakenly think it decides our choices and actions. System 1 is really the one in charge as it “effortlessly originates impressions and feelings that are the main sources of the explicit beliefs and deliberate choices of System 2” (Kahneman *ibid*: 20-21). Most of the time System 1 runs au-

11 The foregoing accounts for the emergence of neuromarketing and its techniques that range from measuring facial expression, skin conductance and pupil dilation to measures of brain activation (cf. Knutson *et al.* (2007 : 147-56); Senior C. and Lee N. (2008 : 263-271); and Ariely D. and Berns G.S. (2010 : 284-92). Likewise, the following articles highlight the shortcomings of traditional marketing tools and approaches: Chartrand *et al.* (2008 : 189-201; Fitzsimons G.J. *et al.* (2002 : 269-279. Another promising, albeit highly controversial, approach is provided by Big Data, with the focus put once more on how consumers actually behave.

12 Nietzsche, 1: “(...) how can man “know himself”? He is a thing obscure and veiled: if the hare have seven skins, man can cast from him seventy times seven, and yet will not be able to say “Here art thou in very truth; this is outer shell no more.”

13 Wilson (2003) calls it the “adaptive unconscious” but Kahneman wanted a complete break from Freud.

14 Typified by the following example: “How many animals of each kind did Moses take into the ark?” The number of people who detect what is wrong with this question is so small that it has been dubbed the “Moses illusion” (Kahneman 2012: 73). By contrast, a very high percentage of respondents in a QCM survey in the United States answered the question “Who was Joan of Ark?” by ticking the “Noah’s wife” box.

tomatically and System 2 is in a comfortable low-effort mode in the background. System 2 is a supporting character who believes him/herself to be the lead actor and often has little idea of what is actually going on.

The fundamental consequence of the dual-process model was summed up by Kahneman (2012: 52) as follows: “The notion that we have limited access to the workings of our minds is difficult to accept because, naturally, it is alien to our experience, but is true: you know far less about yourself than you feel you do”. Wilson (2003: 52) reports that our five senses are taking in eleven million pieces of information every second but at the very most we can process consciously around 40 pieces of information. Accepting that human beings are often unaware of why they acted in a given manner, or believe in blatantly false explanations as to why they did, is a crucial step. We cannot expect such self-knowledge to be directly accessible by researchers. Actually, this explains why positivism fails in social sciences.

Economist Dan Ariely’s book *Predictably Irrational* (2008) in which he explores, *inter alia*, the conflict between social and market norms has a revealing sub-title “The Hidden Forces That Shape Our Decisions”. Studies in this field highlight the astonishing extent to which we are influenced by arbitrary and apparently unrelated factors. The consequences can be dramatic. Kahneman (2012: 43-44) describes a study published in the *Proceedings of the National Academy of Sciences*. Eight judges considering parole applications on average approved 35% of requests. The authors of the study plotted the portion of approved requests against the time since the last food break (morning break, lunch and afternoon break). It spiked after each meal at around 65% and then declined steadily during the two hours or so until the following break, sinking to *about zero* just before the next meal.

3. I can’t get no satisfaction

The marketing industry was all the more fascinated by such findings as it was already aware of the problems encountered in surveys. In 1997 Vavra (1997: 71) came to the conclusion that:

One of the insights in assessing customer satisfaction was the understanding that quality (as delivered in our products and services) is not an objective thing to be measured by conformance to engineering or design specifications. Rather, that *quality (as primary determinant of satisfaction) is, frustrating as it may be to engineers and technicians, a very subjective concept*, depending substantially on individually-derived cues and other soft data. So the first learning for a satisfaction professional is to never assume he or she knows exactly what the customer is looking for, *or how the customer defines quality.* (my italics)

Some twenty years ago, Robert Peterson (1992: 71) had complained that “indeed, examination of the satisfaction literature inevitably culminates in a pervasive yet inescapable conclusion -- *it is not clear what customer satisfaction ratings are measuring* (my italics).”

This very point was in fact raised by Cattaruzza and Mack in 1995 in their survey of simultaneous interpretation users, based on Vuorikoski’s date quality

criteria, since they asked: “Are we sure we were all talking about the same thing?”. Diriker (2011 *passim*) reports that her “semi-structured interviews with users of SI at a philosophy conference (...) hint at the fact that there is not a common and objective understanding regarding quality criteria [...]”. In her book, she drove home this point:

[...] although the respondents seemed to be referring to the “same” quality criteria that were also used by some user surveys (such as “fidelity to the original meaning”, “correct terminology”, “grammaticality”, “fluency”), there were significant differences not only in how users of SI *rated* various quality criteria, but also in how individuals *defined* the criteria. (Diriker 2004: 80)

Unfortunately, the use of surveys runs into many other pitfalls than the straightforward fact that respondents understand key words in different manners.

“Prospect Theory: An Analysis of Decision Under Risk” by Kahneman and Amos Tversky, published in the prestigious *Econometrica* journal in 1979, is one of the most frequently cited articles in social sciences. It showed that alternative decision frames produce systematic changes in responses and choices and even reversals in judgments, despite the fact that the frames are equivalent. For instance, large changes of preferences are sometimes caused by inconsequential variations in the wording of a questionnaire. Likewise, Ofir and Simonson (2001) have shown that forewarning customers prior to service consumption that they will be asked to assess quality afterwards leads customers to report less favourable quality evaluations and reduces their willingness to purchase and recommend the service. This is due to “negativity enhancement”, i.e. forewarned consumers tend to focus primarily on negative aspects of service experiences.

Research in marketing has also shown that significant differences, of around 10-12%, in satisfaction ratings are found between a questionnaire administered orally and a self-administered one, as well as between telephone interviews and those obtained through mail (drawing on “a large, nationally representative sample in excess of 5,000 new car buyers”) (Peterson 1992: 70). Bruine de Bruine (2010: 21) writes that:

[...] on written surveys, options that appear near the beginning of the list are more likely to be selected. [...] It appears that earlier options receive more cognitive processing, with respondents presumably thinking of more reasons for selecting them [...]. Possibly, respondents assume that earlier options are more important to the researchers.

Likewise, the ratings respondents can choose will exert an influence; e.g. whether the option to tick “No opinion” is provided. Such effects apply to all surveys. We now need to look at problems encountered in questionnaires that specifically ask respondents to look back at the last day(s) during which they have been listening to interpreters and judge their performance.

At this point, Kahneman’s “peak-end rule” is highly relevant. He argues that when we look back at a past experience and evaluate it there is a discrepancy between the “experiencing self” and the “reflective self.” We would not be surprised if a friend were to tell us that they were at the opera the night before and had a

fantastic time until the end when somebody's cell phone went off and rang noisily for what seemed an eternity. We would not tick if our friend said that incident had ruined the whole performance. In fact, her memory of the evening may have been ruined but she actually spent hours in musical bliss. Her "reflective self" however is confusing experience with the memory of it, which " [...] is a compelling cognitive illusion – and it is the substitution that makes us believe a past experience can be ruined. The experiencing self does not have a voice" (Kahneman 2012: 381). Kahneman (*ibid*: 388) conducted experiments that confirmed his intuition. "In intuitive evaluation of entire lives as well as brief episodes, peaks and ends matter, but duration does not." Therefore, when delegates fill in a questionnaire at the end of a conference, their reflecting selves are relying on their memory of how the interpreters performed, as determined by the moments when their level of satisfaction (hopefully) peaked and their recollection of the beginning and closing periods. Actually, this might provide an additional explanation to a paradox Gile has mentioned several times (e.g. Gile 1991: 198). Every interpreter has at times come out of the booth thinking "I'm afraid I wasn't very good in this shift" and to her amazement been complimented by users. In fact all parties might be reacting, albeit differently, in accordance with the peak-end rule.

After a conference on Heidegger, a respondent assessing the interpreters she has listened to for one or more days might, for instance, complain about the quality of their work but her memory could be influenced by the fact that she did not understand a lot of the content. What Taleb (2008: 71) calls "narrative fallacy" could be operating here. "Memory is more of a self-serving dynamic revision machine: you remember the last time you remembered the event and, without realizing it, *change the story of every subsequent remembrance*" (his italics). Alternatively, a colleague she admires and who speaks fluent German may praise the interpreters during a coffee break. As a result, whatever the respondent may have felt during the conference could well be overridden and she will voice her satisfaction at the quality of the interpreting.

The foregoing entails that such an innovative survey as the one conducted in 1993 and 1994 by Moser, which included open-ended questions and differentiated between respondents who were listening to SI for the first time and what Moser calls "old timers", different age groups as well as genders, etc., may well provide useful and interesting information. Nonetheless, it cannot hope to measure objectively user satisfaction or quality of interpretation. After discussing why questionnaires are far from problem-free with respect to surveying user expectations and satisfaction, I would now like to focus on two fundamental criteria they make wide use of: fidelity to the original meaning and performance-related criteria.

4. All you do is read in French...

One of the first criteria in the list proposed to users in a questionnaire on the quality of interpreting/user satisfaction tends to be a variant of "sense consistency with the original message", or "fidelity to the original meaning", when it is not the very first parameter.

Diriker (2011: *passim*) demonstrated that, contrary to what was sometimes claimed earlier in the literature, “homogeneity” of the audience could not be taken for granted. Her interviews confirmed that some users listen exclusively to the interpreters while others shifted between the floor and their headset for various reasons. Furthermore, all interpreters know that somebody who is listening to SI for the first time will easily be “wowed” while a frequent user, say in an international organisation, tends to be far more *blasé*. Their expectations are likely to be radically different.

What about users who do not speak the SL? Interpreters must seek to convey the feeling that they are trustworthy by producing a discourse that “sounds” logical and inspires confidence and leaves users with the impression that nothing of importance has been omitted. In particular, when working for live TV an interpreter cannot afford to hesitate or stumble. When discussing SI, Pym (2008: 98) makes a relevant point: “For a mediator of any kind, once you lose trust, you lose everything.” In my opinion, this entails using appropriate terminology but also the “right” jargon given the subject matter, for instance. When interpreting a presentation of financial statements for financial analysts, if the speaker is talking about IFRS-related issues, I would translate the French word “*fusion*” by “business combination.” The analysts would understand perfectly well if I used “merger”, but I hope that unconsciously they are getting the impression that I am familiar with IFRS jargon. Trust is the key word as shown by (Donovan 2002: 4) “participants often choose to listen to the interpretation even if they can “get by” without. They do so for reasons of convenience, *but only as long as the interpreter inspires confidence*” (my italics).

We often have to cope with speakers whose sentences remind me of Jean-Luc Godard’s description of his own films: they have a beginning, a middle part and an end – but not necessarily in that order. In such a case, I will seek to “package” their utterances because I do not want the people listening to me to get the impression that the somewhat incoherent ramblings they would otherwise hear are due to my incompetence. In the same way, one of my rules of thumb is that if the SL people in the audience burst out laughing my listeners should laugh as well. Recently, a speaker I was interpreting disagreed with someone in the room about whether their company had four or three building sites under way in the city of Troyes. To put an end to this rather inconsequential controversy, he quipped “*allez, la guerre de Troyes n’aura pas lieu.*” I translated this brilliant play-of-words by “oh well, you might be right, I was never any good at counting past two.” As the English-speakers had just seen their French colleagues laugh heartily, they were in the right mood and giggled in turn.

As said above, the “fidelity to the original meaning” criterion frequently tops simultaneous interpretation user questionnaires and nearly always is one of the first parameters in the list. It also usually receives the highest percentage in terms of satisfaction, although, as we have seen, its very position could be sending the message that the people who drafted the survey think it is the most important parameter and accordingly its very position may influence answers. This criterion in fact will be frequently interpreted in accordance with a widespread “folk model”, described by Sperber and Wilson (2004: 37) in the following manner:

The speaker's thoughts, encoded into an utterance, should be replicated in the hearer by a decoding process. The result of verbal communication should be an exact reproduction in the hearer of the thoughts the speaker intended to convey.

Consequently, the interpreter/translator ought to be invisible since s/he is merely decoding a message. A striking illustration of this viewpoint, and its implication that translating/interpreting is not all that difficult, is the potential customer who once balked at the rate I charge for a translation and blurted out: "But all you do is read in French and type in English". Language here consists in individuals exchanging messages that have one, and only one, meaning. "Correct" translating or interpreting is expected to consist in deciphering an individual's message and faithfully as well as entirely replicating it in another language. A contrasting viewpoint is defended by Viaggio (2009: 10): "Every act of translation is, at the same time, an act of mediation. The translator's transparency, no matter how desirable in certain contexts, is a myth [...]"

Let me give an example of how different points I have made so far can overlap. At a meeting of an international institution, the head of the French delegation took the floor to voice his frustration because he felt a meaningless discussion was meandering on and on. In fact, in my opinion, he thought a pretext was being used to attack indirectly a key ally of his. He curtly exclaimed "*On ne va quand même pas passer des heures à enculer les mouches.*" The other French delegates laughed somewhat aggressively in agreement. My afore-mentioned rule of thumb kicked in and I said something like "Monsieur X has just complained that this issue has been debated far too long by using an obscene French saying that involves doing unmentionable things to flies." The English-speaking delegates roared with laughter. In the meanwhile, clearly thinking he had gone too far, he looked up at me and added "I suppose you shouldn't translate that". I reverted to the first person and translated his second comment. The French joined in the general merriment as he ruefully sighed "Oh well, too late".

The whole mood of the meeting changed. A tense situation that threatened to deteriorate had been defused. If I had merely interpreted his outburst by saying something tame like "Could we drop this issue, it's been discussed long enough?" in a monotonous tone, the English speakers would have noticed that he had spoken in an abrupt manner and the other French delegates were laughing rather unpleasantly. Moreover, the fact that the Frenchman had requested me not to translate his initial statement also implied he regretted his aggressive tone. Note that I did not mediate consciously, I wish I could react that fast, I was thinking only of the need to get my listeners to laugh.

In a nutshell, the "folk model" discussed above focuses on language as communication between individuals. Language is, by definition, a social activity and is "embedded" in culture and society. In 2011, I attended an event organised by the *Société Française des Traducteurs*. The guest speakers were Ros and Chloe Schwarz. Ros had just published a new translation of Saint-Exupéry's *The Little Prince*. Chloe is Ros's teenage daughter and had been used as a sounding board by her mother since the translation was aimed at people of her age. Ros brought up the thorny problem she had faced in translating Saint-Exupéry's unsavoury phrase "*rois nègres*" and asked for suggestions. Regrettably, quite a few translators indignant-

ly bleated at what they deemed was yet another case of “politically correct” thinking. I tried to point out that “*rois nègres*” was a common, if unpleasant, phrase in French at the time unlike “nigger kings” in English and such a rendering was thus utterly unacceptable. Ros actually liked my proposal, i.e. “tin pot kings”. Unfortunately, when she asked her daughter for her opinion, to my amazement, Chloe said she did not know the phrase. In other words, translating two French words gave rise to a political/cultural problem and an age-related one.

As Sperber and Wilson (2004: *passim*) point out, we can make hypotheses about the speaker’s intentions, and will often be right, but we can never be certain. Diriker (2011: 23) argues that:

[...] receivers cannot access authorial intentions completely because each instance of language use contains more meanings, intentions and accents than its formulator may have intended and any single receiver can purport to have accessed.

When we are working in a booth, we cannot always know when the speaker is making a veiled threat, referring to a scandalous situation, being sarcastic or poking fun at a past statement by someone else, and so forth. As a result, our version in the target language may very well distort or omit a crucial component of the message through no fault of our own since we can hardly know the background information (office politics, etc.) the speaker is referring to, or what people listening to us know, or their grasp of the TL. If I had been interpreting a debate about how to translate “*les rois nègres*” in *The Little Prince*, I would never have realised that most teenagers would not understand the phrase “tin pot kings” (like Chloe, my sixteen-year old daughter did not for instance). In a questionnaire handed out at the end of the debate, an English-speaking teenager may well have given me a negative rating since a key concept had been “badly translated.”

5. ...If it ain’t got that swing

*It makes no difference
If it’s sweet or hot
Just give that rhythm
Everything you’ve got
It don’t mean a thing if it ain’t got that swing
It don’t mean a thing all you got to do is sing
(doo-ah)
Duke Ellington & Irving Mills*

The category of performance-related criteria encompasses such features as rhythm, intonation, fluency, voice quality and accent. Blatant omissions, lengthy gaps in the interpreter’s flow of speech, audible hesitations and sentences that break off abruptly will undermine user trust. Rennert (2010:112-113) adds other flaws such as audible breathing, vowel and consonant lengthening, false starts, repairs, repetitions and speech rate to the reasons that lead the user to get the impression of lack of fluency. She draws the following conclusion:

[...] the results presented above suggest that there is a link between perceived fluency and perception of the interpreter's accuracy, confirming previous studies that suggested that lower fluency may impact negatively on the perceived quality of an interpretation. [...] fluency cannot be ignored as a factor that influences audience perception.

In like manner, Collados Aís (2001: 109) has shown that “monotonous intonation hampers users' retrieval of information” and apparently “has a negative effect on the evaluation of other quality criteria”.

Following in her footsteps, Holub (2010: 117) states: “Analysis showed that monotony can have a negative impact on both comprehension and the assessment of the interpreter's performance.” Yao *et al.* published in 2012 a fascinating article based on Yao's Ph. D. thesis. They scanned the brains of 18 participants using functional magnetic resonance imaging (fMRI) while they listened to audio clips of short stories containing direct or indirect speech quotations. Their main conclusion is that when the brain hears monotonously spoken words it feels should be more expressive, it creates an “inner voice” to drown out the offending speech and the brain simply ‘talks over’ the speech it hears with more vivid speech utterances of its own. Their experiment showed increased brain activity in the ‘voice-selective areas’ of the brain, i.e. certain areas of the auditory cortex. The brain merely needs to be informed that it is dealing with indirect speech to react in this manner. We can assume that listening to somebody with a monotonous intonation for 30 minutes will result in a significant increase in brain activity, and this will be tiresome. For instance, a monotonous intonation means that the speaker's voice does not drop at the end of an affirmative statement. However, listening is a guessing game, we are constantly predicting what will be said next¹⁵, and the listener will expect the sentence to continue. We can easily see the detrimental impact on the user's ability to understand the speaker. Collados Aís *et al.* (2007: 167) report that respondents associate monotonous intonation with the feeling of boredom for the listener. Furthermore, boredom reportedly leads respondents to feel tired and accordingly struggle to understand the content. Moreover, monotonous intonation gives the impression that the interpreter does not like his/her work and/or is not interested in what is being said. I would add that we may assume that the reason why respondents are attending an event is that they are usually interested in the subject and will unconsciously react negatively to their impression that the interpreter finds it excruciatingly tedious.

Collados Aís *et al.* (2007: 97) draw attention to their finding according to which respondents, after ranking content-related criteria above performance-related ones, as is usually the case in such surveys, clearly did not confirm this ranking in their actual assessment of interpreters' performance. They note that users brought up such features as diffidence in the interpreter's voice, shrill tone and monotonous intonation when talking about content-related parameters (*ibid.*: 104). The Pöchlhammer and Zwischenberger (2010) survey corroborates this finding:

15 A key point in Relevance Theory: “The hearer's goal is to construct a hypothesis about the speaker's meaning which satisfies the presumption of relevance conveyed by the utterance.” (Sperber and Wilson, 2004: 619)

The criterion of intonation [...] has an appreciable impact on judgements of overall interpreting quality, even when listening for only one minute. Presumably, the impact of monotonous intonation would be even more pronounced for a standard turn length of up to 30 minutes.

An important conclusion is that *the dichotomy between content- and presentation-related criteria is therefore far less clear-cut than widely believed*. It is confirmed by Diriker (2011: *passim*) “users seem to perceive quality criteria as intertwined constructs with fuzzy borders”. Moreover, the way in which respondents claim they evaluate performance apparently does not match what actually occurs.

In view of the foregoing, it is hardly surprising that Collados Aís (2002: 336) should draw the conclusion that “users not only desire but demand ... a certain degree of intrusion or active involvement on the part of the interpreter”. The interpreter should assume the conscious role of professional communicator and go beyond the ‘ghost role’¹⁶. Granted, any comparison is apples to pears and misleading. I chose a jazz standard as the title of this article because, to some extent, an interpreter is a jazz musician in comparison with, say, a translator – who is more of a classical musician. When jazz musicians are jamming in an inspired manner, despite possibly a few flaws, their overall performance and spontaneous creativity can sweep the audience off its feet. An interpreter who is “on fire” creates a somewhat similar *rapport* with his or her listeners. As Gile (1999c: 159 for instance) has rightly emphasised “the high frequency of errors and omissions that can be observed in interpreting even when no particular technical or other difficulties can be identified in the source speech [...]”. Users will sometimes pick up such flaws, yet nonetheless be satisfied with the interpreters’ performance just like spectators will overlook the shortcomings of a jamming session during a jazz concert because they are delighted by the experience as a whole.

6. The elephant in the room

Although I have criticised them in this article, I would certainly not argue that user satisfaction questionnaires should be discarded, if only because they show that we interpreters are interested in user feedback. Surveys can also undeniably provide useful information as long as we keep in mind their intrinsic flaws. For example, Collados Aís (2001: 109) has recommended, on the basis of the surveys and follow-up interviews she has conducted, that students record themselves interpreting and listen to how they sound in the light of the importance of delivery-related criteria. An interesting point is never discussed in the literature to my knowledge although it deserves looking into. Questionnaires ask respondents to assess the quality of interpreting, although we may safely assume that at the very least two interpreters were involved every time. The implicit and contentious

16 In Collados Aís *et al.* this viewpoint is asserted even more decidedly: “Entre ellos, y a la vista de los resultados, el tema del papel del intérprete como elemento activo y decisivo del proceso de interpretación y experto comunicador, viene a ocupar una posición esencial.” [2007: 223]

message is that interpreters provide homogenous quality. Getting respondents to assess individual interpreters in follow-up interviews and explain differences might actually be useful. In fact, we face the problem of deciding whether we want to carry out objective research or merely seek to defend our professional interests and therefore dodge such controversial issues.

This leads to another crucial issue: all the questionnaire-based quality research in Interpreting Studies I know of breaches a fundamental marketing rule. Surveys poll existing users and most often do not consider former or potential users¹⁷. Questionnaires could be given to every participant in a conference or seminar to ask them whether they are going to use the interpretation and, if not, why? Some former users would presumably say this is because their grasp of the SL has sufficiently improved. But if some delegates have stopped using the service of interpreters because their expectations were not met in the past, we are systematically skewing results. Moreover, once again, follow-up questions as to why they were not satisfied could reveal important information. The same point holds for people who have stopped recruiting interpreters. Is it only because of the expense?

Interestingly, again to my knowledge, SI quality research never mentions sound technicians although corporations and institutions often employ them on a full-time basis. Their viewpoint may well have an impact, albeit an indirect one perhaps, on the decision to hire interpreters or not. It would be interesting to ask them and other people in administrative jobs at agencies, international institutions, Communication Departments and interpreters' secretariats for instance, whether they have any complaints about interpreters and why they prefer to work with some interpreters and not others?

Alain Bonzon, my late friend whom I met 25 years before becoming an interpreter, worked for an international institution and frequently recruited interpreters. I cannot remember him ever criticising interpreters for their work in the booth. On the other hand, when I read an article published by Viaggio in 1996, "*The Tribulations of a Chief Interpreter*", I could literally hear Alain's voice as he indignantly described cases of freelance interpreters, in particular, behaving in a less than satisfactory manner – especially when reading Viaggio's "caricature"¹⁸ of interpreters (he does pointedly comment though: "As any caricature, my appraisal above contains more than a grain of truth"). To be blunt, I have heard the words "divas" and "prima donnas" used in the same sentence as "interpreters" far too often. Most often, this holds for older colleagues, of my age in other words, who regret the loss of status interpreters have suffered from, like many other professions, and is hopefully less true for younger colleagues.

In a nutshell, in contrast with the usual approach, the quality of the service provided by interpreters needs to be assessed in a holistic manner – not just their

17 Peter Moser (1995) is an exception.

18 "When you are a recruiter, it becomes immediately apparent that interpretation is an overpriced [...] service provided somewhat grudgingly by notoriously testy specimens who count minutes the way Scrooge counted gold coins. [...] [Furthermore, they] are complaining every time they do not have a document, or about a speaker going too fast or the slides being projected on the wrong wall [...]."

performance in the booth. Daniel Gile (1995b: 156) is a noteworthy exception since he also mentions the way interpreters dress, behave inside and outside the booth, interact with delegates, etc. I would like the hypothesis I have just outlined, i.e. too often interpreters have reacted negatively to the relative deterioration in our profession's status by behaving in a demanding and/or exasperating manner, to be proven wrong. I would argue, however, that we cannot ignore such a possibility.

7. Completely uninteresting and trivial tautologies

Shlesinger (1997: 129) highlights the fundamental flaw in the idea that accumulating surveys, ultimately, will lead to "scientific" knowledge¹⁹. "However, as long as each questionnaire focuses on different variables, is formulated along different lines, and is administered to different types of target audiences, comparisons of the results will be difficult". We have seen above many other problems that would also rule out such a possibility, first and foremost the fact that the data from such questionnaires are unreliable since respondents' answers cannot be presumed to shed light on their actual thoughts and emotions. Chiaro and Nocella (2005: 172) voice a totally different viewpoint from mine.:

Over and over again we read that investigating quality in interpreting is not an easy task due to the huge number of variables involved [...] The general idea which comes across to the reader is that dealing with the enormous heterogeneity of circumstances in and around interpreting verges on the insurmountable.

They go on to argue:

But is not apparent insuperability typical of scientific enquiry? Was Watson and Crick's model easy to identify? And what of the excogitation of a formula that shows that distance and time are not absolute? And discovering penicillin? The list of seemingly intractable problems is endless. But is it not this very complexity that is what makes research fascinating and irresistible?

Apparent insuperability can also be absolute insuperability, as alchemists eventually discovered. Note also that Chiaro and Nocella mention only examples from so-called "hard" sciences. Such scientism goes hand in hand with another egregious error. The successes enjoyed by research in physics are not accounted for by its methodology *per se*, but the fact that physics adapted its methodology to what it studies, in other words matter. Positivism claims, by contrast, that there is one, and only one, scientific methodology although it concedes that said methodology may be adapted to social sciences. But, to paraphrase French economist Jacques Généreux, I can drop a stone from the top of the Eiffel Tower and calculate precisely where and when it will hit the ground or some unfortunate passer-by. If I throw the same stone at somebody, I cannot predict how they will react, or

19 A viewpoint Thomas Kuhn described as "the textbook image of science".

in fact how I would react tomorrow if somebody threw a stone at me. As Passet (2010: 256) points out, the words “individual” and “atom” share the meaning of “beyond which no further division is possible”. Precisely, positivism postulates that an individual is to society what an atom is to the physical universe. The absurdity of the battle cry of economics, i.e. *ceteris paribus*, consists in the fact that all other things are *never* equal when dealing with social events²⁰. Individuals are not atoms unless you believe that there is no such thing as society. Carr (1960: 31) summed up the argument pithily:

[...] take the dictum of J. S. Mill, the classical individualist: ‘Men are not, when brought together, converted into another kind of substance.’ Of course not. But the fallacy is to suppose that they existed, or had any kind of substance, before being ‘brought together’.

One example springs to mind: the important issue of the role played by interpreters during colonialism or even current wars such as Afghanistan or Iraq. At times, research in interpreting accordingly needs to consider cultural, social and economic factors. Assuming that such research is “unscientific” or less “serious” simply reflects a misunderstanding of what science really means.

We cannot accept the “diktats” of the ATSC and their definition of science, as otherwise Interpreting Research will be doomed to the fate Werner Heisenberg, Nobel laureate for the creation of quantum mechanics, described in *Physics and Beyond*:

The positivists have a simple solution: the world must be divided into that which we can say clearly and the rest, which we had better pass over in silence. But can anyone conceive of a more pointless philosophy, seeing that what we can say clearly amounts to next to nothing? If we omitted all that is unclear we would probably be left with completely uninteresting and trivial tautologies. (1971: 213)

They would presumably include attempts to define quality²¹ “objectively” and debates about the statistical treatment of data generated by respondents’ answers to questionnaires.

To conclude, I would like to emphasize I am not criticising any kind of research *per se*. I am simply arguing that the stifling influence of positivism must be challenged. Asserting that one’s approach is hard-headed and scientific while the so-called Liberal Arts Paradigm is merely woolly-headedness may be gratifying, but bold statements tantamount to dismissing the “Humanities” in fact imply that the proponents of such a viewpoint are somehow entitled to call into question the “intellectual rigour” of fellow researchers who have studied such subjects as philosophy. In fact, first and foremost, we need to move beyond C.P. Snow’s “Two Cultures” mindset.

20 Indeed mainstream economics epitomises all the detrimental consequences of “physics envy”.

21 Such as the “formula” proposed by Kurz (2001: 405): “Quality = Actual Service – Expected Service”

References

- Ariely D. (2008) *Predictably Irrational*, Harper & Collins.
- Ariely D, / Berns G.S. (2010) "Neuromarketing: the hope and hype of neuroimaging in business", *National Review of Neuroscience* 11, 284-92.
- Bruine de Bruine W. (2010) "Framing effects in surveys: how respondents make sense of the questions we ask", in K. Gideon, *Perspectives on Framing*, New York, Psychology Press.
- Carr E.H. (1990/1961) *What is History?*, London, Penguin Books .
- Cattaruzza L. / Mack G. (1995) "User surveys in simultaneous interpretation: a means of learning about quality and/or raising some reasonable doubts," in J. Tommola (ed.), *Topics in Interpreting Research* Turku, University of Turku, 51-68.
- Chartrand T. / Huber J. / Shiv B. / Tanner R. (2008) "Nonconscious goals and consumer choice" *Journal of Consumer Research* 35, 189-201.
- Chiaro D. / Nocella G. (2004) "Interpreters' perception of linguistic and non-linguistic factors affecting quality: a survey through the World Wide Web" *Meta: Translators' Journal*, 49/2, 278-293.
- Chiaro D. / Nocella G. (2005) "Looking quality research in the eye: are we beholding the splinter and ignoring the beam?," *The Interpreters' Newsletter* 13, 167-199.
- Collados Aís A. ([2002/1998) "Quality assessment in simultaneous interpreting: the importance of nonverbal communication" in: F. Pöchhacker /M. Shlesinger (eds) *The Interpreting Studies Reader*, London/New York, Routledge, 326-336.
- Collados Aís A. (2001) "Efectos de la Entonación Monótona sobre la Recuperación de la Información en Receptores de Interpretación Simultánea", *The Translator* 5, 103-110.
- Collados Aís Á. / Macarena Pradas Macías E. / Stévaux E. / García Becerra O. (Eds) (2007) *La Evaluación de la Calidad en Interpretación Simultánea: Parámetros de Incidencia*, Granada, Colección Interlingua Editorial Comares.
- Diriker E. (2004) *De-/Re-Contextualizing. Conference Interpreting: Interpreters in the Ivory Tower?* Amsterdam/Philadelphia, John Benjamins.
- Diriker E. (2011) "User expectation surveys: questioning findings and drawing lessons for interpreter training", <<http://www.iudergi.com/tr/index.php/ceviri/article/viewFile/11090/10348>>.
- Donovan C. (2002) "User expectations" in *EMCI Workshop Proceedings*, 1-13.
- Fitzsimons G.J. / Hutchinson J. W. / Williams P. / Alba J. W. / Chartrand T. L. / Huber J. / Kardes F. R. / Menon G. / Raghuram P. / Russo J. D. / Shiv B. / Tavaoli, N. Y. (2002) "Non-conscious influences on consumer choice", *Marketing Letters* 13: 269-279.
- Gile D. (1990a) "Research proposals for interpreters, in aspects of applied and experimental research on conference interpretation" in L. Gran / C. Taylor (eds), *SSLMIT, Università degli Studi di Trieste*, 226-36.
- Gile D. (1990b) "A communication-oriented analysis of quality in non-literary translation and interpretation" in M. L. Larson (ed.) *Translation, Theory*

- and Practice, *Tension and Interdependence*, Amsterdam/Philadelphia, John Benjamins, 188-200.
- Gile D. (1990c) "Scientific research vs. personal theories in the investigation of interpretation" in L. Gran / C. Taylor (eds): *Aspects of Applied and Experimental Research on Conference Interpretation* Udine, Campanotto Editore, 28-41.
- Gile, D. (1991) "A Communication-Oriented Analysis of Quality," Translation: Theory and Practice". ATA Scholarly Monograph Series (M. L. Larson, ed.), Binghamton (NY), SUNY, vol. 5, 188-200.
- Gile D. (1995a) "Interpretation research: a new impetus?", *Hermes, Journal of Linguistics* 14, 15-29.
- Gile D. (1995b) "La qualité en interprétation de conférence", in *Regards sur la Recherche en Intérpretation de Conférence*. Lille, Presses Universitaires, 143-166.
- Gile D. (1999a) "Use and misuse of the literature in interpreting research", *The Interpreter's Newsletter* 9, 29-43.
- Gile D. (2001) "L'évaluation de la qualité de l'interprétation en cours de formation", *Meta: Translators' Journal* 46/2, 379-393.
- Gile D. (2002) "Conference interpreting as a cognitive management problem" in F. Pöchhacker / M. Shlesinger, London/New York, Routledge, 196-214.
- Gile D. (2004a) "Principles of empirical science: a reminder", <http://www.translationconcepts.org/pdf/Reminder_about_science.pdf>.
- Gile D. (2004b) "Response to the invited papers", in C. Schäffner *Traditions, Gaps and Synergies*, Clevedon, Multilingual Matters Ltd., 124-7.
- Gile D. (2009) "Interpreting studies: a critical view from within", *MonTI. Monografías de Traducción e Interpretación*, 1, Universidad de Alicante 135-155.
- Graves P. (2010) *Consumer.ology*, London, Nicholas Brealey Publishing.
- Grbić N. (2008) "Constructing interpreting quality", *Interpreting* 10 /2, 232-257.
- Heisenberg, W. (1971) *Physics and Beyond*, New York, Harper & Row.
- Holub E. (2010) "Does Intonation matter? The impact of monotony on listener comprehension", *The Interpreters' Newsletter*, 15, 117-126.
- Kahane E. (2000) "Thoughts on the quality of interpretation" <<http://aiic.net/page/197>>.
- Kahneman D. (2012) *Thinking Fast and Slow*, London, Penguin Books,.
- Kahneman D. / Tversky, A. (1979) "Prospect theory: an analysis of decision under risk", *Econometrica* 47/2, 263-291.
- Knutson B./ Rick S./ Wimmer G.E./ Prelec D./ Loewenstein G. (2007) "Neural predictors of purchases"; *Neuron* 53, 147-56.
- Kurz I. (2001) "Conference interpreting: quality in the ears of the user" *Meta: Translators' Journal*, 46/2, 394-409.
- Luccarelli L. and Gree, D. (2007) "Letter from the Editor – standardisation and quality", in *Communicate*, <<http://aiic.net/page/2544>>.
- McCloskey D. (1983) "The rhetoric of economics", *Journal of Economic Literature* 21, 481-517.
- McClosky D. (1990) *If You're So Smart, The Narrative of Economic Expertise*, Chicago, University of Chicago Press.

- Mead P. (2005) "Methodological issues in the study of interpreters' fluency", *The Interpreters' Newsletter*, 13, 39-63.
- Moser, P. (1995) *Survey on Expectations of Users of Conference Interpretation, Final Report*, <<http://aiic.net/page/736/survey-on-expectations-of-users-of-conference-interpretation/lang/1>>.
- Moser-Mercer B. (1994) "Paradigms gained or the art of productive disagreement" in S.Lambert / B.Moser-Mercer (eds) *Bridging the Gap: Empirical Research in Simultaneous Interpretation*, Amsterdam/Philadelphia, John Benjamins, 17-23.
- Nietzsche Friedrich (1871) *Schopenhauer as Educator*, <http://en.wikisource.org/wiki/Schopenhauer_as_Educator>.
- Ofir C. / Simonson, I. (2001) "In search of negative customer feedback: the effect of expecting to evaluate on satisfaction evaluations," *Journal of Marketing Research*, 3, 170-82.
- Passet R. (2010) *Les Grandes Représentations du Monde et de l'Économie*, LLL Les Liens qui Libèrent.
- Peterson, R. A. (1992) "Measuring customer satisfaction: fact and artifact", *Journal of the Academy of Marketing Science*, 20/1, 61-71.
- Pöchhacker F. (2004) "I in TS: on partnership in Translation Studies", in C. Schäffner *Traditions, Gaps and Synergies*, Clevedon, Multilingual Matters Ltd., 104-115.
- Pöchhacker F. / Zwischenberger, C. (2010) "Survey on quality and role: conference interpreters' expectations and self-perceptions", <www.aiic.net/View-Page.cfm/article2510.htm>.
- Pym A. (2008) "On omission in simultaneous interpreting: risk analysis of a hidden effort", in G. Hansen / A. Chesterman / H. Gerzymisch-Arbogast (eds) *Efforts and Models in Interpreting and Translation Research*, , 63-105.
- Rennert S. (2010) "The impact of fluency on the subjective assessment of interpreting quality", *The Interpreters' Newsletter* 15, 101-115.
- Senior C. / Lee, N. (2008) "A manifesto for neuromarketing science", *Journal of Consumer Behaviour* 7, 263-271.
- Shlesinger M. (1997) "Quality in simultaneous interpreting", in Y.Gambier, / D.Gile, / C.Taylor (eds), *Conference Interpreting, Current Trends in Research*, Amsterdam/Philadelphia, John Benjamins, 123-31.
- Sperber D. / Wilson D. (2004) "Outline of Relevance Theory", in L.R. Horn / G.Ward, (eds.), *The Handbook of Pragmatics*, Oxford, Blackwell, 607-632.
- Taleb N. N. (2008) *The Black Swan*, London, Penguin Books..
- Tarran B. (2011) Interview of Rory Sutherland, *Research*, <<http://www.research-live.com/features/measuring-emotion/questioning-the-nature-of-research/4005918.article>>.
- Vavra T. (1997) *Improving Your Measurement of Customer Satisfaction: A Guide to Creating, Conducting, Analyzing, and Reporting Customer Measurement Programs*, ASQ Quality Press.
- Viaggio S. (2004/1996) "The tribulations of a chief interpreter", <<http://aiic.net/page/1324/>>.
- Viaggio S. (2009) "To Translate or to Mediate? That is the Question!", <http://sergioviaggio.com/?cat=7&paged=4>

- Watson J. / Crick F. (1953) "A structure for Deoxyribose Nucleic Acid", *Nature*, 171, 737-738.
- Wilson T. D. (2003) *Strangers to Ourselves: Discovering the Adaptive Unconscious*, Belknap Press of Harvard University Press.
- Yao B. / Belin P. / Scheepers C. (2012) "Brain 'talks over' boring quotes: top-down activation of voice-selective areas while listening to monotonous direct speech quotations", *NeuroImage*, 60/3, 1832-1842.