Thoughts on the quality of interpretation

Is there a consensus on what quality is and how to define and assess it objectively?

Eduardo KAHANE.
Published: May 13, 2000 Last updated: December 2, 2015

AIIC is a professional association upholding the quality of services provided by its members. It has mechanisms to ensure quality now and in the future, and committees to monitor the entry of new members, languages added by current members, training and research. Thanks to its representativeness and reputation for quality, AIIC negotiates agreements with large international organisations on working conditions and remuneration. All this would seem to indicate that there is an underlying consensus on what quality is and how to define and assess it objectively.

Such a consensus would fully justify (to society, clients and users) the primary aim that sets our association apart - quality - as well as the criteria for interpreter training AIIC offers institutions of learning, AIIC´s stance on new forms of interpretation, videoconferencing and interpreting for audio-visual media.

Moreover, we live in an era obsessed with quality control. With that in mind, in 1995 AIIC launched a study on the quality expectations of interpretation users (1).

The amazing thing is that there is no such consensus. Granted, users and interpreters agree on certain quality criteria, but significant differences remain as to nuances, and especially as to the very essence of the elusive concept of quality; quality for whom, assessed in what manner? (2).

I doubt I've been successful in rising to the challenge of addressing these issues here in Communicate (a publication by professional interpreters open to the public at large on the Net). My aim is to draw attention to the problem so as to foster debate, be reasonably thorough in reviewing research to date and avoid off-putting scientific jargon. Here are the results.

The Interpreter's Task

One simplistic definition of the interpreter's task or function is that of mediation to facilitate communication between speaker and listeners. It would be equally simple to add that the simultaneous interpretation (SI) that best achieves that goal would by definition be of the best quality. That may be the ideal model, but in actuality we know things are more complex. The interpreter's mediating function is not clear: is it mediation or interference? We do not even have a clear idea of the speaker-listener relationship. Is it better for the interpreter to integrate into that relationship or to stay outside it? What are speaker-listener relationships like? How do they vary? What is the purpose of the situation wherein communication is taking place? Is it a multitudinous conference or a small meeting of people already well acquainted with each other? What are the mutual intentions and expectations of the participants? Where is or who supposedly is the listener,
The history of Quality Studies, a recent undertaking, bears witness to all these difficulties and may be of use to us to trace the efforts made to overcome them.

**The Beginnings: Quality Expectations**

Initially interpreter-teachers based much on intuition and were occasionally self-complacent toward the "miracle" of interpretation, thus largely blurring early theoretical research on interpretation. This went hand-in-hand with a set of practical rules and precepts for the learning process, such as "interpret ideas not words, finish sentences, etc." Later on the first attempts at analytical models were made.

In 1986, Bühler carried out what was probably the first field study on "quality." She sought to identify and assess the weight of specific factors affecting the quality of SI. She also applied them to professional interpreters, including members of the AIIC Admissions Committee (CACL). Bühler pinpointed as many as 16 criteria. They have the virtue of being the first and in addition have been used in subsequent studies thus enabling a degree of comparability.

Bühler’s idea was to infer users' assessment from that of interpreters. The inference was put to the test for the first time in 1989 by Kurz who used 8 of Bühler’s criteria with users (4). Both studies deal with expectations. A priori assessments are given for generic situations and an ideal interpretation. Comparison of results has a merely indicative value, as follows:

Out of the total sample it was considered as important by:

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Bühler 1986 %</th>
<th>Kurz 1989 %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sense consistency with original message</td>
<td>96</td>
<td>81</td>
</tr>
<tr>
<td>Logical cohesion of utterance</td>
<td>83</td>
<td>72</td>
</tr>
<tr>
<td>Correct grammatical usage</td>
<td>49</td>
<td>45</td>
</tr>
<tr>
<td>Completeness of interpretation</td>
<td>47</td>
<td>36</td>
</tr>
<tr>
<td>Fluency of delivery</td>
<td>49</td>
<td>28</td>
</tr>
<tr>
<td>Correct grammatical usage</td>
<td>48</td>
<td>11</td>
</tr>
<tr>
<td>Native accent</td>
<td>23</td>
<td>11</td>
</tr>
<tr>
<td>Pleasant voice</td>
<td>28</td>
<td>17</td>
</tr>
</tbody>
</table>

The table shows significant percentage differences, but it is worth noting that both groups rank the criteria alike in terms of order of importance. Sense consistency with original message is the most highly valued criterion, followed by logical cohesion. Native accent and pleasant voice are the least valued. Interpreters attach higher value to all the criteria than users do, possibly because they are very self-demanding where quality is concerned, and users have lower expectations. However, interpreters attach significantly greater value to expressive criteria such as native accent, voice quality or correct grammatical usage.

This difference shows up again in later studies. One might ask whether this is an *a priori* value judgement, owing to "knowledge" agreed on in the profession, handed down from generation to generation, or whether it is actually an intuitive assessment, the value of which has gone undetected due to some inadequacy in the research tools. Studies by Shlesinger (5) on presentation and retentiveness and by the AIIC Research Committee point in this direction, highlighting the
importance of prosodic aspects in retentiveness and in user satisfaction.

**Expectations of Distinct Users**

In 1993 (6) Kurz examined the question of distinct user groups possibly having different expectations of interpretation. Her 1989 results were confirmed with no noteworthy exceptions. Likewise those results were confirmed, with some qualifications, by Marrone (7) and Kopczynsky (8).

In 1995 the AIIC Research Committee commissioned Peter Moser to carry out a study of user expectations based on the same hypothesis (different expectations for different types of users). The study was commendable in that it covered 84 conferences, but the results follow the same tendency of greater importance being given to content-related rather than form-related criteria. Even in open questions, theoretical skills and voice are hardly mentioned spontaneously by the subjects of the study. The attempt to determine ideal quality or a set of basic user expectations, regardless of the kind of conference, is hindered by similarity in circumstances.

No matter what the conference topic, new groups of users, with slight methodological variations, are functionally in a similar situation as to practical aspects, group dynamics, and the basic speaker-listener relationship.

Perhaps results would have varied more had studies targeted users in completely different environments - small meetings among participants who all know one another, new product presentations (typically "live", with all sorts of paraphernalia, one or several announcers, use of slogans and timed scripts) or interpreting for audio-visual media.

At last in 1995 Kurz and Pöchhacker (10) were able to match interpreters’ intuition that formal issues and expression are important, with results from a study of users. They looked at a group of Austrian and German television representatives and found that while sense consistency with original message and logical cohesion were still the most highly valued parameters, "TV people put less emphasis on completeness but are particularly sensitive to criteria like voice, accent and fluent delivery".

**Objective assessment of quality**

The material reviewed up until now assesses solely expectations and seeks to establish criteria assumed to be decisive in the quality of interpretation. However, unless these criteria are compared to the results of a real interpretation, it is not possible to determine the extent to which fulfilling the criteria may serve to predict real quality. Pöchhacker (11) is innovative in proposing that we assess users’ cognitive grasp of the message conveyed, measuring variables that may have an impact such as speed, pauses, hesitancy, intonation, fluency, mistakes, register, style, etc. Beyond the methodological difficulties of this approach, even if it were possible to objectively determine and achieve consensus on the real quality of a specific interpretation, said quality might not necessarily be the same as the quality perceived by users or even interpreters.

There are many methodological problems involved. How can we "objectively" measure quality of interpretation in the various situations interpreters come up against? The use of evaluation surveys affords a general idea of the quality "perceived", but the factors influencing that perception remain obscure.

Let us start by acknowledging what scholars agree on: that the criterion on which there is broadest consensus, namely sense consistency with original message, is a hard one for listeners to judge, as they do not know both languages. If the main criterion cannot be assessed by listeners, what is their
basis for judging a given interpretation? Sheer intratextual aspects, or in other words logical cohesion of utterance, which does rank high in terms of expectations? Or do expressive elements, seen a priori as less important, such as fluency of delivery, intonation and voice quality, exercise greater influence than previously thought?

As long ago as 1983 (12), Gile, who throughout his extensive body of work has been tireless in advocating the need for objective, quantifiable criteria, was calling for the creation of a collection of taped speeches and interpretations. The latter could then be evaluated according to criteria mutually agreed upon by research institutions and research could be done on the relationship between objective determinations and subjective evaluations. Even so, Gile himself admits, "information losses, sometimes important in a speech, may be offset by a qualitative transformation of the speech reinforcing its impact." Stenzl shares this observation in the sense that a clear and intelligible text with some information loss may be more useful to the listener than a text seeking to be complete at the expense of clarity and intelligibility. (13).

Collados carried out a controlled laboratory study (14) comparing the assessment of interpretation with monotonous versus "melodious" intonation. Total sense consistency was present in some cases and intentional inconsistencies were introduced in others. Collados found that interpretation with melodious delivery and mistakes was generally rated better than interpretation with a monotonous delivery and total sense consistency.

It may be necessary to qualify the statement that users find it hard to detect discrepancies between the interpreted version and the original in controlled studies or specific speeches. That situation is not comparable to the long sequence of interpreted speeches at a conference where inconsistencies become more apparent as time goes by. Neither is it comparable to question and answer sessions. Picking up on mistakes also depends on how interested the user is in the speech. He or she may discard a great deal of information as irrelevant, which the interpreter is not in a position to do.

**Methodology Issues**

Researchers universally acknowledge the methodological difficulties inherent in studying the elusive concept of quality. There are few tools available aside from evaluation surveys, which focus on different goals and variables, thus making it hard to compare results. Again, circumstances vary from one conference or speech to the next. Researchers want assessment criteria to be isolated and interpretation situations classified systematically and in coordination. Moreover, this seems to be the right way to go in order to reach some sort of agreement on what quality interpretation is, provided agreement is reached someday on what quality is, and quality for whom?

Other obstacles lie in the path of research, such as the fact that some meetings are restricted if not confidential, and professional interpreters are reluctant to have their work closely monitored and compared to that of their colleagues.

**A Broader Perspective of the Field**

Beyond methodological problems, we must clear the path by breaking away from universal definitions of "quality" in ideal conditions. Let us classify types of meetings and interpreting situations as this might lead to a series of models that are closer to actual practice. It also seems necessary to broaden the field by moving from purely linguistic issues to pragmatic, communication issues.

Queries about the pleasantness of a voice seem inadequate as a criterion to assess what is known in media circles as voice personality and communication credibility. Voice elements include timbre and
tone, and voice dynamics span such factors as will, energy and the ability to connect with others. A broader perception of the interpreter's mission not just as communication mediator but as a communicator in his or her own right, responsibly taking on a certain role, could open up new paths for research into objective quality and the perception of said quality. A consideration of the legitimacy of this attitude, claimed by Gile (15), with all that it entails, including the acknowledgement that being faithful to the original does not necessarily mean rendering the message in toto, is not only ethically valid, but could have an impact on our image.

Knowing the target audience and their expectations should also be a factor in research, and in a professional's store of resources. Not only do speakers have a different attitude from listeners (the former tolerate a greater degree of "intervention" from interpreters to correct minor errors; the latter seem to prefer literal respect for the speakers' words and even his mistakes), but different listeners in the same situation may have different expectations. Take the actual case of an international delegation, primarily made up of judges, on a visit to a penitentiary. The magistrates posed questions couched in lofty terms the prisoners were not used to. The interpreter toned down his version to make it more accessible to the inmates. The delegation was not unanimous as to what sort of interpretation was preferable under the circumstances. The split was along cultural lines. The Northern Europeans were in favour of literalism; the Mediterraneans supported "interpretation". Viaggio reports a similar situation at the UN where the Chinese delegation favours literalism over fluency or style (16).

The target audience will also include other interpreters when they must rely on a "pivot" to do their work. Although there is consensus that "relay" should be avoided to the extent possible, it is actually used more frequently than we care to admit. Relay is standard procedure at the JICS (the European Commission's Joint Interpretation-Conferences Service which covers some 12,000 meetings a year) for 11-language meetings, or even meetings with fewer languages, and is well on its way to become institutionalised with the coming enlargement. Since the issue has not yet been studied, we still do not know whether the "ideal" interpretation for listeners is equally ideal for interpreters who use it to produce yet another version.

Is it better for the "pivot" version to be very literal or to be recreated in the target language? Are some language pairs preferable to others? What role does cultural proximity play in this triangle? How does the interpreter working with relay extrapolate from the original? How to make up for the loss in immediacy? What expressive qualities are preferable? How to juggle the different needs of users and interpreters taking relay? No doubt this subject deserves greater attention, yet I am aware of only two papers and a lecture devoted to it (17).

We have already pointed out that the media has different expectations regarding interpretation. In the media, interaction with the listener is impossible and the interpreter's performance is judged according to the usual standards of expressive quality for announcers or reporters, not to mention the special traits of the program: narrative, descriptive, theme-based, and so on.

It remains to be seen what the expectations of the users of new technologies, such as video and teleconferencing, will be. The sense of alienation and coldness in communication experienced by interpreters who have tried the new technology may be shared by users. At this point it is imperative to ascertain those expectations and design means to offset the technical and emotional difficulties in the new working conditions.

**Non-linguistic, non-expressive matters**

On occasion, as Gile points out (18), the perception of quality has to do with the interpreter's physical appearance and ability to integrate, or actually to go unnoticed, within the user group.
Discrete and professional behaviour in and out of the booth contribute to how the interpreter's service is assessed. Our professional image is enhanced if we take care to avoid shuffling papers, pouring water, clearing our throats or coughing into an open microphone to ensure users receive clear sound. The same can be said for a knowledge of protocol, taken for granted for consecutive interpretation at high level meetings, although such training is not usually part of standard programs at interpreting schools.

All the concepts described so far and the goal of defining what quality interpretation is, seem closer to the ideal than the practical. Interpreters face trials daily: poor sound quality, lack of documents, speed of speeches, difficulty seeing slides or speakers, non-native accents, private jokes, dubiously pronounced phrases borrowed from other languages without warning, etc. Nevertheless, the perception of quality is affected by these vicissitudes that are hard to explain to non-interpreters and inevitably sound like a poor excuse.

There are, however, virtually no studies assessing objective and perceived differences in performance taking into account factors such as whether the interpreter had documents and time to prepare them beforehand, whether speeches were read or improvised, etc.

Two years ago, JICS, the largest interpretation user, launched its own survey about the quality of the service it offers, adding cost and management considerations to strictu-sensu considerations of quality with the idea that quality interpretation is also dependent on organisational inputs such as programming, engagements, staff selection, etc. They understand that the service offered requires perfect operability, control of the situation on the ground and final assessment. This study, still in progress, will include a typology of the most frequent kinds of meetings, which together with other factors and hypotheses, will be fed into a matrix.

**Conclusions**

While there is a certain level of agreement on some of the factors that are decisive for the expected quality of interpretation, such as sense consistency between the original and the interpreted version, it is not clear which factors influence the perception of that quality, since it is difficult to design universal tests to determine quality perception objectively and users do not know the departure language. Expressive factors could influence perception.

The lack of consensus among interpreters and among users as to the concept of quality itself makes the task harder. Given the many situations where interpretation is used, it would be advisable to establish a thorough classification or typology, including interpretation for the media, the use of relays, etc. Methodological difficulties point to the need for multicenter studies and agreement on criteria decisive for quality to facilitate comparability.

It might be enriching to open up the field to pragmatic issues and include elements of information theory, the communication situation, or even group dynamics. New forms of interpretation and interpretation for the media could be used to broaden the field of research.

The development of a wider body of research, or the mere opening of a debate on quality would shed light on this key issue, which until now has been obscured by a consensus that recognises its importance but lacks substance. Pursuing the issue would provide us with new criteria for both the exercise and teaching of the profession, and for enhancing the way our work is perceived by users and society at large.

---

**References**


Translated by Ann Goslin

Recommended citation format: