# Evaluation, Language, and Untranslatables

**Authors:**
Peter Dahler-Larsen[1] with Tineke Abma[2], Marı´a Bustelo[3], Roxana Irimia[4], Sonja Kosunen[5], Iryna Kravchuk[6], Elena Minina[7], Christina Segerholm[8], Eneida Shiroma[9], Nicoletta Stame[10], and Charlie Kabanga Tshali[11]


1 University of Copenhagen, Copenhagen, Denmark
2 Department of Medical Humanities, VU University Medical Center, Amsterdam, the Netherlands
3 Complutense University of Madrid, Madrid, Spain
4 Pluriconsult Ltd., Bucharest, Romania
5 University of Helsinki, Helsinki, Finland
6 National Academy of Public Administration, Kyiv, Ukraine
7 Institute of Education, Higher School of Economics, Moscow, Russia
8 Department of Education, Umea° University, Umea°, Sweden
9 Universidade Federal de Santa Catarina, Floriano´ polis, Brazil
10 University of Roma, Rome, Italy
11 Independent Consultant International Development, London, United Kingdom


**Corresponding Author:**
Peter Dahler-Larsen, University of Copenhagen, Øster Farimagsgade 5, Copenhagen, 1353, Denmark.
Email: pdl@ifs.ku.dk

**Abstract**
The issue of translatability is pressing in international evaluation, in global transfer of evaluative instruments, in comparative performance management, and in culturally responsive evaluation. Terms that are never fully understood, digested, or accepted may continue to influence issues, problems, and social interactions in and around and after evaluations. Their meanings can be imposed or reinvented. Untranslatable terms are not just ''lost in translation'' but may produce overflows that do not go away. The purpose of this article is to increase attention to the issue of translatability in evaluation by means of specific exemplars. We provide a short dictionary of such exemplars delivered by evaluators, consultants, and teachers who work across a variety of contexts. We conclude with a few recommendations: highlight frictions in translatability by deliberately circulating and discussing words of relevance that appear to be ''foreign''; increase the language skills of evaluators; and make research on frictions in translation an articulate part of the agenda for research on evaluation

**Prologue**

Have you heard about the evaluator who tried to explain the meaning of evaluative inquiry to a Spanish-speaking audience? In their language, they heard her talking about inquisition. Since I recently came across a book with the interesting title: Dictionary of Untranslatables: A Philosophical Lexicon (Cassin, 2014), I have been thinking about untranslatables in evaluation.

In an era of ever more globalized evaluation discourse, it is too often assumed that key terms in evaluation are directly and smoothly translatable into any language. However, can terms such as efficiency, outcome, indicator, performance, achievement, participation, empowerment, and program theory—and inquiry!—be translated? Do they retain their meanings across contexts? In a similar vein, are there key terms in local languages that tend to become forgotten, ignored, or deprived of meaning because they are too difficult to translate into official and recognized evaluation language?

The issue of translatability is pressing in international evaluation, in global transfer of evaluative instruments, in comparative performance management, and in culturally responsive evaluation. Translatability or lack hereof is a part of the work of any evaluator who conceives of herself or himself as a broker and facilitator of understanding across various structures of meaning, whether or not these structures are distinct enough to be recognized as formal languages or they are vernaculars that characterize socioeconomic or ethnic groups or subcultures and whether or not the evaluative endeavor aims mostly at learning, critical dialogue, or deliberation. Translation is in fact a broad phenomenon that manifests itself both substantially and metaphorically when boundaries are crossed, when different kinds of knowing interact, when rules meet practices, and whenever uneven phenomena must be made commensurable in human interaction (Freeman, 2009). While the role of language has been given some attention in evaluation (Agar, 2000; Patton, 2000), and the importance of language issues has been noted in culturally responsive evaluation (Hood, Hopson, & Kirkhart, 2015), more deserves to be said about the friction involved in translation. The purpose of this article is to increase attention to the issue of translatability in evaluation by means of specific exemplars provided by people involved in evaluation in various national and cultural contexts. You will find a microlexicon of untranslatables in evaluation in this article. The entries have been provided by a diverse selection of international evaluators, teachers, consultants, and translators. They have been asked to provide examples that best illustrated, according to their experience as boundary crossers, the most intense and interesting issues with translation in evaluation.

**Translation as a Theoretical Problem**

More often than not, we translate a term by "looking it up" in a dictionary to find an equivalent term that we understand. There are a number of theoretical reasons, however, why there may be a lack of equivalence between terms in different languages. According to the so-called Sapir–Whorf hypothesis, languages condition our cognition, and since languages are fundamentally different, members of different languages live in fundamentally different worlds. Untranslatability rather than translatability is the normal condition according to a strong interpretation of the Sapir–Whorf hypothesis. A more moderate version gains support from cognitive research, showing that some phenomena can be grasped in similar (though not necessarily universal) ways among members of language communities with different terminologies for the phenomena at hand (Kay & Kempton, 1984).

Since de Saussure, structuralist and poststructuralist theories of language have suggested that a term derives its meaning not from any inherent essence therein but from its relations to other terms in a larger system of meanings. Translation is imperfect since it implies the transplantation of term into a new set of relations to other terms.

The pragmatic tradition claims that a term derives its meaning not from the language system itself but from its pragmatic use in social interaction and communication. The key to meaning lies not inside a term but in its social life: Who would want to use that term for which purposes? What are the practical conditions that make the use of that term possible? Content is not the only thing that matters. Style and genre are also important in practice. This adds to translation problems.

The world is not equally transparent from all points of view. When a particular reality becomes linguistically defined and stabilized in organizational procedures and more or less institutionalized, it also defines what counts as an argument and a meaningful statement. Therefore, Community A is sometimes in position to define a set of terms in which Community B must express itself but can't (Lyotard, 1988). Untranslatability is related to power, whether it is recognized or not.

Disagreements occur not only over words but over concepts. Concepts are special words loaded with sociopolitical meanings and expectations embedded in changing historical contexts (Koselleck, 2004). A newly translated word can be ''tamed'' into an established discourse, but it can also, in some circumstances, give voice to a new set of expectations with transformative social consequences.

A constructivist view emphasizes how language, interaction, and imagination help constitute reality (Castoriadis, 1997). Meanings emerge through definition, interaction, and ratification that are collectively embedded (Taylor, 2016). That is also why evaluation of some fairly insignificant phenomenon can become socially significant over time, as evaluation instruments constitute and define what they are claimed to measure (Dahler-Larsen, 2014; Hansson, 2000), and expectations about the use of evaluation set social expectations in motion. Some terms that sound similar before and after translation are in fact false friends, because the meaning of the new term is more narrow, specific, and productive in a particular direction than the untranslated term or alternative terms that could have been chosen. See for example "impatto/impact" below.

Contemporary theoretical perspectives on translation acknowledge that translation is active and productive interpretation. Translators are more than servants to original authors. They are, in fact, more like writers (Freeman, 2009). Rather than the frequently used "lost in translation", found in translation is sometimes a more appropriate metaphor.

**Translation Issues in Evaluation**

For the sake of simplicity, let us talk about evaluator languages and local languages. Evaluators speak varieties of evaluator language or "Evaluese" which may include terms such as achievement, indicator, stakeholder, complexity, outcome, responsive, model, intervention, and many others. Even if evaluators evidently draw from different parts of its terminology and even if empirical studies show that some of its key terms are unexpectedly ambiguous or confusing (Christie, 2003), it still makes sense to talk about "evaluator language" because a large part of the theory, knowledge, tools, and even skills of evaluators is more or less codified into something that can be talked about, used in evaluation models, and applied to structure evaluation processes.

Evaluator language is often a form of English, but it does not have to be. The official language of *American Journal of Evaluation* is English. English is also used frequently enough in international settings to be statistically dominant. English has historically been the language out of which many key terms in evaluation theory have been carved. However, these facts cannot be abstracted from the sociopolitical reasons for the strong role of English in the global arena. Although the relation between evaluator language and local language can be analytically separated from the problem of English versus non-English languages, people involved in issues of translation (such as the authors of this article) do in practice not have the luxury to separate the two kinds of problems.

Some may think of evaluator language as a lingua franca that makes interaction possible between various groups (as did Latin in science, Italian in banking, and French in diplomacy in some centuries). A lingua franca is a trade language that is voluntarily adopted by various groups for functional reasons. While evaluator language sometimes plays that role, I believe that the frictions between evaluator language and local languages are often deeper and more serious.

By "local languages", I mean the terms in which people involved in evaluation and affected by evaluation spontaneously express their lived experience and their view of the world (Schwandt & Burgon, 2006). The notion of "lived experience" suggests that the relevance structures that characterize most people in practice may have to do with handling the practical life they are thrown into, taking into account who they are and whom they want to be. It can also be stated as a brute fact that most local languages were not granted permission to be constructed on condition that they first consulted a list of categories that are equivalent to key terms in evaluator language. In fact, a language such as Danish has no native terms directly equivalent to achievement, stakeholder, or performance.

Translation in evaluation is a two-way street. There can be translation issues from evaluator languages to local languages and from local languages to evaluator languages. (To complicate matters further, some translations involve a third language, such as Russian in Ukraine and French in Democratic Republic of the Congo [DRC]).

Issues of translatability can evidently relate to not only outcomes, criteria, and the like but to evaluation itself (see *arviointi* and *ociniuvannia* below) and all its aspects, including evaluands, evaluation criteria, evaluation models, evaluation processes, values, methods, and forms of use. Evaluators and their social roles are also linguistically typified (see *qallunak* and *mjuaji* below). How participants are characterized and characterize themselves and are represented politically and linguistically is a whole issue in itself.

Not surprisingly, some untranslatabilities also refer to the context in which evaluation takes place. A rough categorization: Some phenomena are structural and formal, such as legal regulations and institutional rules; some are cultural and informal, having to do with tacit values and worldviews; some refer to particular sociopolitical problems in the situation at hand. The message inherent in terms, such as distorção idade-série, multirepententes, and progressão continuada (see below), is that since terms are important in the local context, there is a need for an evaluator to learn more about the sociopolitical context in Brazil to carry out school evaluation in that context.

Some terms have an uncomplicated literal meaning but are not easy to use because it would require much local sensitivity to find out how to operationalize them in evaluative practice (e.g., *pårørende* and *avnämare*). Finally, some terms around evaluation make sense to me only if I understand a particular cultural mentality or philosophical inclination (see *trage vragen* and *Bildung* below).

One problem has to do with understanding. We usually claim to "understand" something when the meanings of a phenomenon can be grasped from our perspective. There is (with Gadamer) a "fusion of horizons", so that what I can grasp overlaps with what is meaningful from the perspective of the Other. This overlap can result not only from the efforts of the Other to explain herself or himself but also from my efforts to reach out and extend my horizon of meanings. Understanding in that sense requires good will and active work. Sometimes we cannot understand because we do not want to stretch out to understand what appears to us as weird, strange, or horrific.

If we take power into account, however, some meaning systems are put in better institutionalized positions than others. Some vocabularies are mandatory. Sometimes it is taken for granted that one group must explain what it is up to in terms of the vocabulary defined by another group (Becker, 1996). Depending on the social, political, institutional, and organizational affiliation of an evaluator and a program person, the two have different obligations to reach out, understand, and explain things (or not) when the evaluator says "tell me about the intended outcome of your program". Evaluators, program people, funders, and others proceed even if horizons of meaning do not overlap. Lack of linguistic competence is not considered a problem by all.

Therefore, the practical consequences of translation problems do not stop if there is a lack of understanding. Terms that are never fully understood, digested, or accepted may continue to influence issues, problems, and social interactions in and around and after evaluations. Their meanings can be imposed or reinvented. Untranslatable terms are not just lost in translation but may produce overflows that do not go away.

Examples of such overflow occur when the problem of untranslatability is "solved" simply by importing a foreign term into a local language. For example, the term "outcome" is now being used directly in Danish. Presumably, this is indicative of a change of policy frameworks in the direction of a more "outcome-oriented" form of public management, but the term itself is also instrumental in that change, not only a symptom.

Sometimes the creation of an Anglicist neologism is deemed appropriate simply because no fitting appropriate term exists in the language in focus. Thus, in Ukraine, the terms validyzaciya and tranguliaciya have been coined to cover validation and triangulation. But such approach may develop a professional jargon

that is not understandable for wider publics and nonexperts. A "purist" translation may keep overflows under control but may also be more difficult to sell in practice.

Concerns like these are considered by people who seek to develop systematically a new vocabulary of evaluation in a particular linguistic context[1]. In other situations, perhaps more frequently, untranslatability is dealt with pragmatically on an ad hoc basis, without much reflection. Sensemaking then grows out of earlier choices in an organic and complex way.

We will only know about translation issues if we consult people who actually work as boundary crossers. The following entries have been selected because they illustrate the most pressing issues in translation as seen from the professional and personal perspective of each of the authors of this article. Each entry contains a message and a learning opportunity for an evaluator. We know the term is important in the local context because people embedded in local translation processes have selected each entry carefully for your consideration.

In the face of the enormous task of how to sort exemplars into some conceptual superstructure, not to mention how they can be captured empirically and re-represented, I shall simply follow organizing principle of the book that inspired this article (Cassin, 2014): All terms are explained in alphabetical order with cross-references.

**Exemplars**

*Arviointi (Finnish):* The equivalent to assessment and evaluation as concepts. Various formulations and footnotes are used to enhance the different meanings of the same word. Päivi Atjonen (2007, p. 20) refers to arviointi as evaluation in the sense of large-scale evaluation of things, such as evaluations of schools or educational policy, or to arviointi as assessment in the sense of individualized pupil assessment and related to grading (which in Finnish is translated into arvostelu, which in English can also be retranslated into judging). Also other terms such as appraisal and monitoring can be added to the list of words that in Finnish are translated into arviointi (Varjo, Simola, & Rinne, 2016, pp. 15–16). Even peer review is translated into vertaisarviointi (peer-arviointi; Jakku-Sihvonen & Heinonen, 2001).

Newer terms are also translated into some kind of Finnish or rather "Finglish", such as evaluation as evaluaatio, monitoring as monitorointi, auditing as auditointi (with stronger reference to practices in accounting than social sciences as a theoretical field; Jakku-Sihvonen & Heinonen, 2001), and benchmark

evaluation into benchmark-arviointi (Jakku-Sihvonen & Heinonen, 2001). The question then remains, if these terms are actually understandable in Finnish as such or if their purpose is mainly to clarify the reference to the English counterpart, as all of the cases above could be translated into plain arviointi with footnotes. Interestingly enough, even if Finland is scoring well in international learning-outcome rankings, the Finnish vocabulary of discussing different forms of assessment and evaluation remains very limited. Finland does not conduct national-level standardized testing, provide school rankings for comprehensive schools, nor implement school inspections. The focus on evaluating, assessing, and inspecting the system and its actors has been minor. Simola, Rinne, Varjo, Pitkänen, and Kauko (2009, p. 174) suggest that the Finns have been "rather effective in resisting a trans-national policy of testing and ranking" and that there is even a silent consensus about antipathy toward ranking.

*Avnämare (Swedish; Plural: avnämare; Antonym, in English: provider, producer, manufacturer, supplier, salesman):* An individual or a group (of), organization(s), employer(s), customer(s), or citizen(s) who receive, use, benefit, or buy a product and/or (public) service. The term is similar to stakeholders but does not include individuals or groups that have no direct use, benefit, or buyer/ employer relation to what is evaluated.Acommon example of how the term is used in evaluations and quality assurance activities is to talk about how the perspective of "avnämare" should be granted. By avnämare in relation to higher education are usually meant the future employers of students, but by avnämare in relation to public social service are usually meant the clients or beneficiaries. However, what avnämare is in relation to a particular evaluation is not always easy to pinpoint, so the choice of a relevant set of avnämare is a practical decision made under the circumstances.

*Bildung (German):* Bildung is an idea(l) of moral development or moral cultivation. There is no one-to-one translation of the word Bildung in English, nor in other European languages. The meaning of Bildung can best be captured by describing its long history and underlying ideas. Bildung includes a personal process of being formed (gebilded werden) after some higher normative ideal or horizon (Bild, e.g., God). This is not an instrumental process of goal attainment, but rather a praxis. The philosopher Hans-Georg Gadamer (1975) argued that dialogical processes of understanding lead to Bildung. Gadamer stands in a long tradition of philosophers who argued that Bildung should be central to academic education. Von Humboldt's 19th-century idea(l), revived in the 1960s by Jürgen Habermas, was that universities take upon themselves the task to promote civilization and make world citizens out of their students with open minds, not afraid to judge and think for themselves. Gadamer considers venturing into the position of others through dialogue to be crucial to this process. He even defines Bildung as "trained receptivity to otherness" and argues that not thinking about one's own standpoints, concerns, and interests for once, but rather making an empathetic and honest attempt to understand the other, brings about a transformation of the self through which one

cultivates oneself. Bildung has thus clear moral (and originally theological and humanist) undertones; it is a process related to questions about the good, either what entails a good life or what it means to be, for example, a good doctor or teacher. Such moral questions start in practice and can be answered in dialogue with others. The mentioned philosophers have therefore argued that dialogue needs to be at the center of our academic praxis, to bring about moral learning processes and a joint and moral understanding of what, for instance, good teaching and research entails. Some evaluation scholars and traditions put dialogue at the center of their praxis to promote Bildung.

Distorção idade-série *(Portuguese as spoken in Brazil):* Literally meaning age/grade distortion. There is an official expectation that children start school at age 6 and finish grammar school, called Ensino Fundamental, 9 years later. However, the expected age/grade relation is not met by 20% of basic school students (and 30% in rural areas). This is because some pupils start school when they are older than 6 and because some leave school to work and sustain their families. In some schools, a student repeats a school year if he or she has not learned the content by the end of the year (see multirepetentes). Since the age/grade distortion is now a component in a national quality indicator, it is significant as evaluative information. Some schools react by promoting students automatically to score better on the indicator (see also *Progressão Continuada*).

*Evaluación responsiva (Spanish):* For lack of a direct translation into Spanish, responsive evaluation (Stake, 2004) was first translated into Spanish as evaluación sensible (literally "sensitive evaluation") and later into evaluación comprensiva (that has two meanings "understanding evaluation" and "comprehensive or extensive evaluation"), but none of these terms were to precisely convey the meaning of the original. Using the official Spanish dictionary, it was discovered that the term responsivo/responsiva actually exists in Spanish (with the proper meaning), although it is rarely used nor intuitively understood. With the risk of being seen as an anglicism, the term evaluación responsiva is now in circulation, accompanied with explanations of how the term should be understood. This example illustrates that sometimes the road to direct translation is not direct.

*Impatto (Italian):* Originally: Clash, collision, as when there is a car accident; the act of hitting something. It has also the metaphorical meaning of influence, for example, an event or a person that has influence on an audience. In general, it is linked to the idea of something violent and negative.

In the Italian evaluation language, impatto has been introduced with Valutazione di impatto ambientale, or environmental impact assessment, where interventions having a high impact are subject to changes and further regulations. Here, high impact means something affecting negatively the environment, with pollution, desertification, and so on.

This evaluator language is not aligned with the current practice of, and debate on, *Valutazione di impatto* or impact evaluation. As is well known, there are two main definitions of impact evaluation (Stern et al., 2012). A limited, methodological one that having defined impact as ". . . the difference in the indicator of interest (Y) with the intervention (Y1) and without the intervention (Y0)" states that "an impact evaluation is a study which tackles the issue of attribution by identifying the counterfactual value of Y (Y0) in a rigorous manner". (White, 2010, p. 154). An enlarged, content based, definition of impact evaluation, based on the Organization for Economic Cooperation and Development/Development Assistance Committee lexicon, states that it is the evaluation of ". . . positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended".

In both cases, having a high impact is an instance of the success of a program. At the moment, in academic circles and even in institutional ones, impact evaluation in the limited meaning is holding the scene in Italy, and it seems that the twist from a negative association to a positive one (did the forecasted effect of the cause take place?) has gone unnoticed. In this context, impatto is identified with a specific theory of causality (by difference, in the counterfactual logic) and keeps apart all what speaks of openness, unintended consequences, societal forces, and a generative causality

*Mcunguzi (or mkaguzi; Kiswahili):* An inspector or investigator. In Eastern Congo (DRC), very few people speak French (the official language in addition to four national languages), and evaluators may have difficulties in explaining their role. They may go around the problem by referring to themselves as mcunguzi, but it has a different connotation, meaning someone who is coming to do policing rather than someone who can judge and assess the merit or the worth of something.

*Mjuaji (Kiswahili; Plural: Wajuaji):* The term comes from the verb ''kujua,'' meaning to know in Kiswahili. ''Mjuaji'' can be translated as a knowledgeable person or an expert. In eastern Congo (DRC), mjuaji can also have a negative connotation when it refers to someone who wants to show off their knowledge. Such a person is not willing to learn from or listen to others; they do not have any humility. Usually, a mjuaji does not get any respect as they tend to annoy others by displaying their pretended knowledge, while in reality others think they are ignorant. In a community, people would not associate a mjuaji to decisions about important matters. When outsiders, especially Western people come to a community to do things without understanding the local context and culture and don't make an effort to learn about it, they get labeled mjuaji or "wajuaji". Labeling someone, or a group of people, or an organization as mjuaji, or wajuaji, affects how people relate to them. People's attitude toward a mjuaji is not usually a challenging one but a passive one, they will say to themselves "leave him/her until when he/she realizes that she/he did not really know as

much as he/she thought.'' If an evaluator has been labeled as a mjuaji, he or she may not get much cooperation from program participants or program staffs, as they will not open much to him or her. This situation is more likely to happen when using nonparticipatory approaches to evaluation or programming.

*Multirepententes (Portuguese as spoken in Brazil):* Students who have failed at school multiple times and repeated the same grade several times. Some of them drop out. Some finish with a diploma but have a hard time at the labor market because they lack fundamental skills. Some return to school later to complete primary or secondary schooling attending "youth and adult education". The term exemplifies how a socioeconomic issue is embedded in an evaluative information such as indicators (see distorção idade-série).

*Ociniuvannia (Ukrainian):* The term chosen by the Ukrainian Evaluation Association as the best translation of evaluation. In order to develop a common national evaluation vocabulary, it was deemed most feasible to build on existing Ukrainian terms. The choice was between ocinka and ociniuvannia. Whereas ocinka refers to school grades, final results, reports, and conclusions, ociniuvannia comes closer to a process of valuing, a process with certain requirements, procedures, and stages.

*Plekken der moeite (Dutch):* Coined by the philosopher Harry Kunneman. Literally this can be translated as "places which are difficult". The word "plekken", however, is different from the word places; plekken refers to the emotional, embodied experiences of situations. Moeite (hebben met) is referring to a situation which is emotionally hard and painful for the participant(s). It is not so much a cognitive difficulty which is at stake, but rather a moral or relational one. Furthermore, moeite (doen voor) includes an activity, some labor that needs to be done to handle the situation. Thus, plekken der moeite denote tensed situations where people confront the boundaries of their problemsolving capabilities. They may experience ambiguity, that is, a vacuum of meaning or an overflow of meaning, along with feelings of confusion and ambivalence. Often language falls short to make sense of the situation. People tend to avoid these plekken der moeite, because they might lead to a loss of safety and control. Kunneman (2009) considers the entrance of plekken der moeite a central part of learning processes[2]. Entering plekken der moeite offers possibilities to handle different perspectives and to create room for change. Reflection on existing patterns in thinking and action and interaction rules may create communicative spaces. Communicative spaces, a notion Kunneman borrows from Jürgen Habermas, offers possibilities to develop new perspectives on situations. The notion plekken der moeite is important for evaluators who place learning and development at the center of their praxis.

*Pårørende (Danish; Plural: Pårørende):* A pårørende is someone who is related to a patient or client though not necessarily through kinship. Pårørende are expected to be informed (though some rules apply concerning

confidentiality) and expected to care about the person in focus. Since their relation with the person in focus is personal and affective, and since Danish language has no term for stakeholders, they are not regarded as stakeholders, but they are often included as relevant partners in evaluation of health care and social services, especially if the person in focus is unable to articulate his or her own views due to dementia or serious illness. Sometimes pårørende engage collectively in larger issues (such as a patient union) in which case they perform the role of what is known as stakeholders in the international literature.

*Progressão Continuada (Portuguese as spoken in Brazil*): Literally meaning continuing progression. It refers to a legal demand (in some states and municipal public school systems) to secure automatic promotion from one grade to the next regardless of the achievements of the individual pupil. The purpose is to reduce the age/grade distortion (see *distorção idade-série)* and cut costs. The term exemplifies why evaluators need to inform themselves about legal frameworks around policies and programs under evaluation.

*Qallunak (Greenlandic; Plural: Qallunaat):* The lexical meaning of Qallunak is "a Dane". Among the Inuit (a self-description of indigenous Greenlanders that merely translates into "human beings"), Qallunak was used to denote all strangers of non-inuit origin. Further meanings of the term flourish among the Inuit of Nunavut, Canada, whose language is similar, but not identical to the Greenlanders'. Here, Qallunak means "White man" (typically modern, Western, etc.). Minnie Aodla Freeman describes in her book "Life Among the Qallunaat" (Freeman, 2015) further characteristics of the Inuit experience with the Qallunak. Qallunaks do not respect the local circumstances, especially weather conditions. They think they know the answer to everything. They are strangely obsessed with time and money. They behave as if life is a burden. They almost never smile. And they have no idea about how to party. These are the connotations you are likely to be met with if you come to Greenland as an evaluator under a category commonly translated, according to the lexical definition, as a Dane.

*Rezultat (Romanian):* A term for results that has been used in a situation where the terms outputs and outcomes were confusing, because they had no equivalents in Romanian. Since Romanian is related to languages with Latin roots, a French vocabulary was later seen as the most useful tool to establish differences between different kinds of rezultat.

*Standart (Russian):* The neoliberal notion of outcome-based standardization, introduced into Russian system of education in the course of post-Soviet modernization reform, is based on the idea of "educational standard" as a "principle of educational provision and governance", aimed at ensuring fair distribution of educational resources and unifying educational content. The learner-centered and competency-based educational standards were positioned in the official state rhetoric in a larger humanistic paradigm as a

"social contract" between an individual, the society, and the state, with learners' developmental needs proclaimed to be of supreme value. In pedagogical terms, the concept emphasized individuality, creative independent thinking, and competency building.

The perception of standardization in the public discourse, however, draws on distinctly different interpretative frames (Minina, 2014). Educational standard is seen in the public mind as a mechanism of exercising state control over education as well as an accountability requirement put in place by the state to dominate educational institutions. It is often referred to with terms such as "corral", "boxed-in", "muzzles on academic freedom", and "burden for teachers". While the official rhetoric emphasizes equalizing educational opportunities and unifying educational content for all, in the public mind, the term "standardization" is reinterpreted as averaging out student achievement on the basis of the lowest acceptable quality. Conceptualized in terms of a manufacturing standard, standardization reform is appraised in extremely negative and judgment-laden Russian terms, including "educational McDonald's", "uravnilovka" (averaging out, depersonalization), "vseh pod odnu grebenku" (literally: "to groom everyone with the same comb", "one size fits all"), and "shtampovka" ("assembly line" or "cut and dry" production).

The backbone of those ideas is the long-standing pedagogical tradition of fostering "nonstandardness" ("nestandartnost") in education and resisting the "gray uniformity" of Soviet-era schooling. The nestandartnost in broader philosophical terms is understood as "oneness", in the sense of the individual uniqueness of each human being, while the "nonstandard" (adj.) is "one" or "one-ofa-kind". Nonstandardness signifies "one-ness", or "equality within individuality", while "standardness" means "sameness", "same as everyone", "stereotypical", "mediocre", and "equally depersonalized". This interpretation of nonstandard is based on the idea of cooperative problemsolving through creative (nonstandard) tasks (nestandartnie zadachi), resulting in independent (nonstandard) thinking (nestandartnoie myshlenie). Standardization, in turn, is seen in pedagogical terms as knowledge centerdness, rationality, and outcome, in which the sole purpose of education is to transmit the ready-made sociocultural heritage of adults to the younger generation.

*Trage vragen. (Dutch):* Coined by the philosopher Harry Kunneman. Translates literally as "slow questions". "Vragen" means questions. The word "traag" is, however, more complicated. In Dutch, a distinction is made between "langzaam" and traag. Both refer to some sort of slowness. Langzaam is related – in terms of meanings – to clock time. This is measurable time. Traag, on the other hand, is not measurable and related to the personal experience of time, and to one's personal rhythms. Langzaam and traag reflect different time perspectives, which cannot be found in the word slow. Traag has existential layers of meaning. The description trage vragen does therefore not refer to questions being asked in a slow manner. Metaphorically,

trage vragen denote questions that cannot answered immediately, the answering of these questions takes time because of their existential and moral nature. The confrontation with trage vragen typically occurs in situations of death and loss, and this often leads to uncertainty and confusion over one's identity and (the continuation of) one's life narrative. Trage vragen beg for the acknowledgment of powerlessness and loss. They need to be digested and ''doorgewerkt'' (not accepted, but actively worked upon, worked through) by a person. Trage vragen require interpretations that reckon the raw nature of reality. The harsh reality should not be ignored or idealized but requires emphatic understanding among other people. This understanding may then create trust to handle the openness and uncontrollable nature of life and form a potential source of personal development and growth. *Trage vragen* cannot be measured and require qualitative methods and responsive designs to address them. The recognition of trage vragen serves as a general warning against an instrumental approach to evaluation.

*Verksamhet (Swedish):* A collection of activities in an organization (company, enterprise, public sector) fulfilling a certain aim; the (daily) work of an individual; work directed to certain ends (in general). The term is broader that "activity" and therefore often includes several types of separate activities; practical, ethical, intellectual, and so on, aimed at a common end. "Verksamhet" is often used to denote the work collectively performed in an organization. In ordinary language, it could be used as follows: "In our verksamhet we take great care to enhance the development of each individual child". Or: "In my evaluation verksamhet I need to be very clear about what the power relations among the stake holders are, before I sign a contract". What verksamhet precisely denotes is specified through the relation to its context.

*Virkningsevaluering (Danish):* When a version of theory-based evaluation was to be introduced in Denmark, the term theory was not found to be helpful in introducing this model to practitioners. Realistic evaluation was also considered, but not chosen, as it was feared that some practitioners would question whether alternative evaluations were presented as unrealistic. Virkningsevaluering was coined as a term that captured "virkning" which refers to both the process and the result of how things "work". Etymologically, the Vikings brought the roots of this term into the English language. While the choice of virkningsevaluering as a neologism perhaps helped overstate the originality of the contribution, it allowed Danish evaluators to define this model for their own purposes, and it allowed researchers and students to trace the specific use of the introduced model over time.

**Epilogue**

Perhaps you have already thought, as I did, that the *Dictionary of Untranslatables* is a paradox, because it spends 1,297 pages carefully explaining the meaning of hundreds of terms across languages, which would be

futile if the untranslatable were beyond the comprehensible. To make the paradox complete, the whole book has been *translated* from the French into English.

Admittedly, as the preface (p. xiv) says, untranslatability is never absolute, but there can be more or less imperfect translations. In a very small scale, this article has tried to do the same. Efforts to communicate the untranslatable rest on the assumption that since we are civilized, democratic, and intelligent, our efforts to understand phenomena from the perspective of meaning systems other than our own will bear fruit. Sapir and Whorf, we do not live in cultural and linguistic closed boxes. Untranslatables are not essentialist predicates of nation, ethnos, or culture with no equivalent in another language, but reminders of singularities of expression in a worldscape full of semantic dissonance (Apter, 2014, p. xv). We do live in a world where interaction is common and necessary, and we must found out how to deal with the tensions, ambiguities, and confusions of the apparently untranslatable. Translating is better conceived as an ongoing social process than as transportation and replacement of terms.

How we deal with translation issues in the field of evaluation may be more important than how we translate terms. If we already know that process use is important in evaluation, we should realize that translation as a process is central for the construction of meanings, relations, purposes, and consequences related to evaluation.

Too often, however, problems related to translation become invisible for us. Paradoxically, we think we understand the Other when we can express a term, an issue, an intention, or meaning in the world of the Other in terms our own language. Then translation is already complete (Hacking, 2002). Or rather it appears to be complete from our perspective, but not necessarily from the perspective of others. Issues of translation are embedded in larger institutional systems that make some languages stronger and some weaker (Asad, 1986). The general warning, I wish to issue is to not assume that evaluator language is the lingua franca that makes these differences irrelevant. Instead, any language, including evaluator language, is, for better and for worse, an interpretation, an intervention, an interference, and a construction.

If you already speak English as a part of your evaluative practice in multilingual situations, translation effects may be invisible to you when you have the privilege of having all evaluative communication translated into your language but that does not mean that translation problems do not exist.

To pay attention to the untranslatable is to create a space for the anguish, the concern, and the hesitation experienced when there is not enough room for what we as human beings do not want to see translated, because we would miss the meanings we believe are original as we would miss a friend or a child (Apter,

2014, p. xiv). What the notion of untranslatability does is to enhance our sensitivity to the friction in translation. It reminds us that in order to understand we must make an effort.

What can be done, more specifically? We can highlight frictions in translatability by deliberately circulating and discussing words of relevance that appear to be "foreign", thereby signifying peculiar differences in meaning (and in doing so, we should do more than just exhibit the exotic for the tourist; the exotic should be more than names of exotic drinks and world music rhythms in pop songs. By displaying exemplars, this is surely a risk that this article runs).

We can increase the language skills of evaluators. We can reflect upon the limitations of our own language, including evaluator language.

We can make research on the frictions in translation an articulate part of the agenda for research on evaluation. As the *Dictionary of Untranslatables* suggests, we can make a virtue of seeing the differences (Apter, 2014, p. xiv).

**Declaration of Conflicting Interests**

**Funding**

**Notes**

1. For an example from Ukraine, see http://www.ukreval.org/images/Glossary.pdf

2. The work of Harry Kunneman is mainly in Dutch and has unfortunately not been translated in English. This

is understandable as his work contains so many ideas which are hardly translatable.

**References**

Agar, M. (2000). Border lessons: Linguistic "rich points" and evaluative understanding. *New Directions for Evaluation, 86,* 93–109.

Apter, E. (2014). Preface. In C. B. Cassin (Ed.), *Dictionary of untranslatables. A philosophical Lexicon* (pp. vii–xvi). Princeton, NJ:  Princeton University Press.

Asad, T. (1986). The concept of cultural translation in British social anthropology. In J. Clifford & G.E.Marcus (Eds.), Writing culture (pp. 141–164). Berkeley: University of California Press

Atjonen, P. (2007). *Hyvä, paha arviointi.* Helsinki, Finland: Tammi.

Becker, H. S. (1996). The epistemology of qualitative research. In R. Jessor, A. Colby, & R. Shweder (Eds.), *Ethnography and human development: Context and meaning in social inquiry* (pp. 53–72). Chicago, IL: University of Chicago Press

Cassin, B. (Ed.). (2014). *Dictionary of untranslatables. A philosophical Lexicon*. Princeton, NJ: Princeton University Press.

Castoriadis, C. (1997). The imaginary: Creation in the social-historical domain. In C. Castoriadis (Ed.), *World in fragments* (pp. 3–18). Stanford, CA: Stanford University Press.

Christie, C. (2003). The language of evaluation theory: Insights gained from an empirical study of evaluation theory and practice. The Canadian Journal of Program Evaluation, 18, 33–45.

Dahler-Larsen, P. (2014). Constitutive effects of performance indicators: Getting beyond unintended consequences. *Public Management Review*, 16, 969–986.

Freeman, M. A. (2015). Life among the qallunaat. Manitoba, Canada: University of Manitoba Press.

Freeman, R. (2009). What is ''translation''? *Evidence and Policy*, 5, 429–447

Gadamer, G. H. (1975). Wahrheit und Methode [Truth and method]. Grundzüge einer philosophischen Hermeneutik, in: Gesammelte Werke (Vol 1). Mohr: Tübingen.

Hacking, I. (2002). *Historical ontology.* Cambridge, MA: Harvard University Press.

Hansson, F. A. (2000). How tests create what they are intended to measure. In A. Filer (Ed.), *Assessment. Social practice and social* product (pp. 67–82). New York, NY: Routledge.

Hood, S., Hopson, R., & Kirkhart, K. (2015). Culturally responsive evaluation. In K. E. Newcomer, H. P. Hatry,& J. S. Wholey (Eds.), *Handbook of practical program evaluation* (4th ed., pp. 281–318). Hoboken, NJ:Wiley Online Library.

Jakku-Sihvonen, R., & Heinonen, S. (2001). *Johdatus koulutuksen uudistuvaan arviointikulttuuriin.*Helsinki, Finland: Opetushallitus.

Kay, P., & Kempton, W. (1984). What is the Sapir-Whorf hypothesis. *American Anthropologist*, 86, 65–79.

Koselleck, R. (2004*). Futures past. On the semantics of historical time*. NewYork,NY: ColumbiaUniversity Press.

Kunneman, H. (2009). *Voorbij het dikke-ik.* Amsterdam, the Netherlands: Boom.

Lyotard, J. F. (1988). *Le Différend [Phrases in dispute].* Minneapolis, MN: University of Minnesota Press.

Minina, E. (2014). 'Why doesn't the telephone ring? Reform of educational standards in Russia'*. Inter Disciplines: Journal of History and Sociology*, 5, 1–44. Retrieved from http://www.inter-disciplines.org/index.php/indi/article/view/124

Patton, M. Q. (2000). Overview: Language matters. *New Directions for Evaluation*, *86*, 5–16.

Schwandt, T. A., & Burgon, H. (2006). Evaluation and the study of lived experience. In I. Shaw, J. Greene, & M. Mark (Eds.), *The Sage handbook of evaluation* (pp. 98–117). London, England: Sage.

Simola, H., Rinne, R., Varjo, J., Pitkänen, H., & Kauko, J. (2009). Quality assurance and evaluation (QAE) in Finnish compulsory schooling: a national model or just unintended effects of radical decentralisation? *Journal of Education Policy*, *24*, 163–178.

Stake, R. (2004). *Standards-based and responsive evaluation.* Thousand Oaks, CA: Sage.

Stern, E., Stame, N.,Mayne, J., Forss, K., Davies, R., & Befani, B. (2012). Broadening the range of designs andmethods for impact evaluations. Report of a study commissioned by the Department for International Development, London, UK.

Taylor, C. (2016). The language animal. Cambridge, MA: Harvard University Press.

Varjo, J., Simola, H., & Rinne, R. (2016). *Arvioida ja hallita. Perään katsomisesta informaatio-ohjaukseen suomalaisessa koulutuspolitiikassa* [To evaluate and to govern: From looking-after to information steering in Finnish school politics]. Jyväskylä, Finland: FERA.

White, H. (2010). A contribution to current debates in impact evaluation. *Evaluation*, 16, 153–164.