

Seppo Kittilä*

General knowledge as an evidential category

<https://doi.org/10.1515/ling-2019-0027>

Abstract: Numerous studies have dealt with the coding of information sources within and across languages. These studies have shown that despite the significant differences in the number of formally distinct evidentiality categories, languages tend to have grammaticalized markers for certain information sources, but not for others; different kinds of sensory perception, inference, assumption and hearsay evidence are among those information sources that receive explicit coding. In this paper, another evidence type, namely general knowledge, will be examined. It will be shown that general knowledge differs from other information sources in its nature, but it also has features in common with them. In addition, a formal-functional typology of general knowledge coding will be proposed based on the nature of the element used for this purpose. Finally, the rationale behind the discussed types and the central theoretical implications of the paper will be discussed. The attested types either stress the peculiar nature of general knowledge providing evidence for its independent information source status, or they emphasize the common features shared by general knowledge and other information sources, most notably reliability of information.

Keywords: evidentiality, general knowledge, information source, typology

1 Introduction

Evidentiality as a linguistic category refers to the source of information speakers have for their statements. Evidentiality can either be a grammaticalized, obligatorily expressed category, as, e. g. in Wutun, Tariana (see (1)) and Hup, or languages may refer to the information source optionally by lexical means, such as lexical verbs or particles, which is the case, for example, in, English, German and Finnish (2). Consider:

*Corresponding author: **Seppo Kittilä**, Department of Modern languages (General Linguistics), University of Helsinki, P.O.Box 24, Helsinki 00014, Finland, E-mail: seppo.kittila@helsinki.fi

- (1) Tariana
- a. *Juse irida di-manika-ka*
 Jose football 3SG.NF-play-REC.P.VIS
 ‘Jose has played football (we saw it)’
 - b. *Juse irida di-manika-mahka*
 Jose football 3SG.NF-play-REC.P.NONVIS
 ‘Jose has played football (we heard it)’
 - c. *Juse irida di-manika-nihka*
 Jose football 3SG.NF-play-REC.P.INFR
 ‘Jose has played football (we infer it from visual evidence)’
 - d. *Juse irida di-manika-sika*
 Jose football 3SG.NF-play-REC.P.ASSUM
 ‘Jose has played football (we assume this on the basis of what we already know)’
 - e. *Juse irida di-manika-pidaka*
 Jose football 3SG.NF-play-REC.P.REP
 ‘Jose has played football (we were told)’
 (Aikhenvald 2004: 2-3)
- (2) Finnish
- a. *Jaakko on kuulemma fiksu*
 Jaakko COP.PRES.3SG hearsay smart
 ‘They say that Jaakko is smart’
 - b. *Näen, että Aino piirtää hyvin*
 see.1SG.PRES that Aino draw.PRES.3SG well
 ‘I see that Aino is drawing well’
 (personal knowledge)

Tariana is a language with a rather elaborate evidentiality system with five grammatically marked categories, one of which is always marked on the verb. In Finnish, evidentiality is expressed by (grammatically) fully optional particles, or lexical verbs that express meanings similar to some Tariana evidentiality markers. For example, in (2a), the hearsay particle is optional formally, even though its use may be of the utmost importance pragmatically.

Earlier studies of evidentiality have shown that languages differ according to what kind(s) of information source they code grammatically and how many categories they distinguish explicitly (see, e. g. Willett 1988; Plungian 2010; and especially; Aikhenvald 2004: Ch. 2). Despite the evident differences between languages in this respect, previous studies have shown that the coded types of information source are rather constant across languages, even though languages

naturally vary according to how many distinctions they make. A list could look, e. g. as below (proceeding from the most direct to the least direct evidence):

1. Ego-evidence
2. Visual evidence
3. Other sensory evidence (auditive, tactile, olfactory, gustatory)
4. Inference
5. Assumption
6. Quotation
7. Hearsay

The list above is based on the information sources usually discussed in studies of evidentiality (the typology above is my own, even though it is based largely on Aikhenvald (2004: 63–64), somewhat different typologies are found, for example, in Willett 1988; and Plungian 2010 discussed below). The list thus provides us with a rather exhaustive picture of how humans gather information about the surrounding world and what kinds of information sources we use for our statements and what kind of information sources receive linguistic coding (see also the discussion of Plungian below). First of all, we are personally involved in numerous states-of-affairs every day, in which case our own (volitional) participation in a given event serves as our information source, as in *I am working on this paper* or *tomorrow I will eat fish for dinner*. This corresponds to the definition of ego-evidence in this paper; ego-evidence refers to the speaker's own (usually volitional) participation in the event denoted (the evidence type has also been labeled as participatory evidence or egophoricity in later studies). It is, however, important to note that ego-evidence is not always seen as a genuine information source, and, for example, Aikhenvald (2004) does not consider it in her study. In this paper, ego-evidence/egophoricity is taken into account, because it is a formally distinct category in languages such as Wutun and Tibetan (see also, e. g. Floyd et al. (2018), for more studies), and it is also directly relevant to general knowledge coding. Moreover, we make numerous statements based on our own actions, as we do based on, e. g. visual evidence or hearsay, which further speaks in favor of considering ego-evidence as an information source in its own right.

Second, sensory perception is an important way of gathering information about the world; we constantly witness states-of-affairs by seeing, hearing, feeling, touching and smelling. Our senses differ according to their importance and reliability, and it is no surprise that many languages give visual information special status treating it as the most direct (and reliable) information source. Some languages also have an auditive evidential used for auditory sensory evidence (see, e. g. Jalava 2017 for Tundra Nenets), but no language has a

distinct evidential for olfactory, gustatory or tactile evidence. In addition to observing events directly, we can infer causal relations between different events (inference) and/or assume something less directly based, e. g. on our knowledge of the world (assumption). Inference presents more direct evidence than assumption, since inference is based on less controversial evidence leaving less room for speculation (Barnes (1984: 262) notes that in Tuyuca assumptive evidential is used if the speaker has no evidence for his/her claim). Finally, we may acquire information from others, in which case we are dealing with hearsay or quotative information. In the first case, the information source is not accounted for, while the information source is known for quotation. In English, this corresponds roughly to the differences between *they say that John is sick* (hearsay) and *according to Lisa, John is sick* (quotation).

Even though the examined information sources are above presented as independent, it is important to note that the sources may overlap, and they have features in common. This is manifested, for example, in the somewhat different kind of taxonomy of information sources proposed by Plungian (2010: 37; see also Willett 1988):

Direct/personal (=attested, witnessed, firsthand, confirmative)

- Participatory/endophoric; common knowledge
- Visual (with subtypes)
- Non-visual (sensory)

Indirect/personal

- Inferential (based on observed results)
- Presumptive (based on plausible reasoning) (common knowledge)

Indirect/non-personal (secondhand)

- Reported (with subtypes)

As such, the information sources in Plungian's typology are largely the same as those proposed earlier. However, Plungian classifies the discussed information sources based on their (in)directness and (non-)personal nature. It therefore accounts better for the relations between information sources than the list above. A crucial distinction is made based on whether the information is our own or not; on the list above, types 1–3 represent direct information, while 4–7 correspond to indirect information on Plungian's list. This is an important distinction, because we can, for example, usually take responsibility for statements based on our own evidence, while we cannot do this for information we receive from others (see also, e.g. Aikhenvald 2002: 190). This difference is relevant to the discussion in this paper, because general knowledge constitutes the speaker's personal information.

In this paper, another kind of information source that has received less attention in previous studies will be discussed. More precisely, the present paper is concerned with general knowledge as an evidence type. The coding of general knowledge (or facts depending on the source) is occasionally noted (see, e. g. Faller 2002a; Aikhenvald 2004; Loughnane 2009; Plungian 2010; Gawne and Hill 2017a), and the use of factual evidentials has been studied rather extensively in Tibetan languages (see e. g. Oisel 2013; Gawne and Hill 2017b for studies of this), but to the best of my knowledge, cross-linguistic, systematic in-depth studies focusing exclusively on general knowledge are lacking to date. The primary goal of this paper is to show that semantically general knowledge constitutes an evidence type of its own despite its partial similarity with other instances of direct evidence. General knowledge comprises here pieces of information that have become the speaker's internal information, of which the speaker has absolute subjective certainty, and for which s/he does not need any kind of external evidence (see Section 2 for a more detailed definition). General knowledge is thus viewed at least to some extent in the same way as in the classification proposed by Plungian; it is seen as the speaker's own (endophoric) evidence (common knowledge in the first sense). General knowledge is harder to classify based on one feature only, because, as will be discussed in Section 4, general knowledge shares features with other information sources, but its nature differs clearly from them in other respects. Here general knowledge is primarily viewed via reliability; it is viewed as reliable evidence, since it is the speaker's own (endophoric) evidence the speaker has absolute certainty of. This follows, in addition to other reasons discussed in detail in Section 2, since general knowledge is in many languages coded by direct evidentials.

In addition to a semantic definition of general knowledge, its formal coding along with the rationale behind this will be discussed. Most languages do not code general knowledge with a formally distinct marker but use some of the existing evidentiality markers for this (usually a marker of direct evidence). This means that we often do not have formal evidence for the independent nature of general knowledge. However, as will be shown, the use of the employed markers is not random, but languages display clear tendencies in this respect. In other words, the features general knowledge has in common with other information sources are manifested in its linguistic coding. Finally, the relevance of the discussion in this paper to our general understanding of evidentiality will be discussed.

A few methodological/terminological remarks are in order before proceeding. First of all, it is important to note that the states-of-affairs briefly described above could be labeled as both facts and general knowledge. Neither term is completely without problems, but in this paper, I have opted for using the term

general knowledge. This follows primarily, because general knowledge accounts for the type of evidence discussed better. The label fact is associated too much with certainty, which means that also events that we are witnessing ourselves as we speak, or our own inner states could be seen as facts, since we “know them for a fact”. However, these states-of-affairs are not discussed in this paper, because they do not constitute general knowledge, which also speaks in favor of choosing the term general knowledge for the type of evidence examined here. Second, it is important to note that even though the present paper discusses data from various languages, it is not a genuine typological paper in that its findings are not based on a carefully selected sample. The goals of the paper are thus primarily theoretical in nature. The main reason for the unsystematic nature of data collection is that the linguistic coding of general knowledge is not always discussed in the consulted sources. Third, the discussion of general knowledge in Section 2 concerns only basic cases, i. e. indisputable pieces of general knowledge. Pieces of information that have just become known to the speaker and whose information status is thus somewhat different, will not be discussed, even though they are occasionally used as a basis of comparison. Following from this, the typology proposed in Section 3 is based on clear cases as well.

The organization of the paper is as follows. In Section 2, semantics of general knowledge will be discussed including both similarities with and differences from other information sources. Section 3 proposes a formal-functional typology of general knowledge coding based on the functional nature of the element used for this purpose. Section 4 examines the rationale behind the attested types, and finally Section 5 summarizes the central findings of the paper.

2 Semantics of general knowledge

2.1 Preliminaries

In this section, I will define in detail what I mean by general knowledge. This also includes a discussion of the most important differences from along with certain similarities between general knowledge and other information sources, i. e. what renders something general knowledge, and how these pieces of information differ from other information sources will be scrutinized. The goal of this section is to show that semantically general knowledge constitutes an information source in its own right. It is also in order to note that the speaker usually

has a reason for coding something as general knowledge, s/he may, e. g. wish to stress the reliability of the given claim, even though another kind of coding would be possible. However, the discussion below will be centered round the semantics of general knowledge instead of pragmatics.

General knowledge is in this paper defined as pieces of information that fulfill the following criteria:

1. General knowledge is a part of the speaker's established world view (even though it is originally based on external evidence). It constitutes the speaker's internal information, and the speaker has subjective certainty of the truth value of the given information.
2. The speaker can refer to general knowledge without any kind of external evidence (sensory/hearsay evidence, inference, assumption). This also makes general knowledge more stable and less time-dependent than other sources of information, since no external evidence is needed. We can freely choose when we refer to general knowledge.
3. General knowledge is based on the speaker's previous experiences of the world, but the original source of information (which can be of any of the types discussed above) does not need to (or even cannot) be specified.

According to the definition above, *2 + 2 is 4* and *Stockholm is the capital of Sweden* constitute paradigm cases of general knowledge, since these are both generally known facts about the world. Also states-of-affairs such as *John and Lisa have two children* and *Lisa is a professor of mathematics* may equally well be general knowledge to those who know these states-of-affairs based on their previous experiences. On the other hand, *John is shopping at the moment* is not general knowledge, because we need some kind of external evidence for this statement, and the statement holds true only as long as John is shopping. It is important to note that basically the same piece of information can be general knowledge or, e. g. visually evidenced event, as is exemplified, e. g. by *Lisa and John are jogging in the park*. In case we see Lisa and John jogging in the park for the first time, we are dealing with a statement based on visual evidence, and not general knowledge. On the other hand, if we know that John and Lisa jog regularly in the park, and not, e. g. in a forest, we are dealing with general knowledge. Whether we are dealing with general knowledge or not is thus not determined solely by the nature of events. The nature of evidence changes when a piece of information becomes general knowledge. In the second case, we can make the statement *Lisa and John are jogging in the park* regardless of time and place, while in the first case we need evidence for our claim (see also Zemp 2017:

268–269 for the use of factual and direct testimony markers in Purik Tibetan). Below, I will discuss the three features of general knowledge in more detail.

2.2 Subjective certainty

In principle, all states-of-affairs we have been involved in, and for which we thus have some kind of evidence potentially constitute general knowledge. However, only a very small percentage of all possible states-of-affairs are genuinely general knowledge for us. This follows, because general knowledge is a part of our established world view, and this applies only to those states-of-affairs we find relevant enough. For example, if we have a Facebook account, we are informed of friends' birthdays regularly, but we usually rather quickly forget who was born when, which renders the use of external evidence necessary. Only the birthdays of our closest friends usually become general knowledge for us. Our own internal world is different from that of all other speakers, which has the consequence that people differ drastically as to what they may code as general knowledge. Frequency of occurrence and habituality are highly relevant in this regard; states-of-affairs that occur frequently in our immediate environment easily become general knowledge for us. This is formally manifest, for example, in Turkish, where the non-firsthand evidential no longer appears with reported information, when the given piece of information has become a part of the speaker's general knowledge (Aksu-Koç and Slobin 1986). Similar cases have been reported, e. g. for Yukaghir (Maslova 2003: 229) and Mamainde (Eberhard 2012).

As noted above, the speaker has subjective certainty of his/her general knowledge. In other words, the speaker knows a certain state-of-affairs to hold even though s/he has no direct evidence for it as s/he speaks. For example, we may have subjective certainty of the fact that Stockholm is the capital of Sweden even though we have no concrete evidence for this as we speak. Subjectivity has the, sometimes unfortunate, consequence that we may believe something to be general knowledge (and a fact), even though this is not true objectively. For example, someone may think that the capital of Australia is Sydney, even though this "fact" does not reflect the actual state-of-affairs in the world. The objective truth value is thus sometimes secondary, but the speaker's own subjective certainty determines whether s/he chooses to code something as general knowledge. It is, however, important to note that the linguistic coding of general knowledge is not random, and the speaker must believe that his/her own beliefs coincide with the actual factuality of information. For example, we do not (if we act rationally) view something like *two plus two is seventeen* as general

knowledge/fact, because we know this to be false, and we thus cannot have subjective certainty of it. Related to this, we may also forget things that have been general knowledge for us. For example, if we do not need chemistry after school, we will probably forget most of the periodic table. We may, e. g. need to rely on some external evidence for the symbol of Tellurium, in which case this is no longer general knowledge for us. In other words, general knowledge is something we have absolute subjective certainty of at the moment when we refer to a given state of affairs.

Finally, the subjective nature of general knowledge is related to the way in which the speaker chooses to refer to a certain piece of information. Representing something as general knowledge is usually a deliberate choice with a clear communicative function (other means may also be available, see, e. g. Eberhard 2012 for Mamainde; Yliniemi 2017 for the use of the different copulas in Denjongke). For example, the speaker may wish to underline his/her certainty of a specific piece of information, or s/he may wish to communicate that no other kind of evidence is needed, and/or s/he is no longer aware of the original source. In general, the speaker does not refer to a piece of information as general knowledge (or fact) in case s/he cannot take responsibility for his/her claim.

2.3 Lack of external evidence

The internal nature of general knowledge has the very natural consequence that we do not need external evidence for referring to these pieces of information. The only kind of evidence we need, and can have, for general knowledge is our own internal, subjective certainty about the denoted state of affairs (see also Stenzel 2008: 425 for Wanano, where the assertion evidential used also for fact coding is used whenever the speaker has no external evidence). Referring to any other kind of evidence deprives a piece of information of its status as general knowledge. Lack of external evidence constitutes a very clear difference to other information sources; we need visual evidence for using the visual evidential, while no evidence is needed for general knowledge. This also drastically changes the nature of the evidence, and thus underlines the special nature of general knowledge.

The lack of external evidence makes general knowledge independent of time and place; we can freely choose when and where to use our internal evidence as a basis for a statement. In other words, we do not need to wait for the right moment for stating *Gaborone is the capital of Botswana*, but we can do this whenever this is relevant. For all other types of evidence, we have to rely on

external evidence, whose availability we cannot determine. We thus cannot say *John is coming* or *it is raining* whenever we wish (if we really mean what we say), but this is possible only under the right circumstances. This is also relevant to the two readings of *Lisa and John are jogging in the park* noted above (claims based on visual evidence are time- and place-dependent). Moreover, general knowledge applies as long as we have no evidence to the contrary. Put concretely, evidence for general knowledge is the same at moment 1 and moment 2, while for other states-of-affairs the nature of the evidence may vary depending on the context (e. g. it may rain at moment 1, while it does not rain at moment 2). Certain pieces of general knowledge belong to the past, such as *Bonn was the capital of West Germany*, but as long as the truth value of a piece of information is constant, and no external evidence is needed, it constitutes general knowledge for the speaker.

Lack of external evidence and the closely related independence of time and place make an important contribution to what can become general knowledge for us. Probably the paradigm case of general knowledge is illustrated by scientific facts, such as *2 + 2 is 4* and *the Earth orbits the Sun*. These facts apply regardless of time and place, and (at least in principle) we cannot have disconfirming evidence for them, which makes the differences between external and internal evidence irrelevant. Second, states-of-affairs such as *John and Lisa have two children* and *Lisa is an English teacher* are also rather constant in nature. Third, events that occur habitually in our immediate environment often become internalized. For example, if we see Lisa arrive at her office every day wearing a biking helmet, we may deduce *Lisa comes to work by bike* to hold generally. The least likely candidates for general knowledge are illustrated by states-of-affairs in constant flux. Examples include, e. g. the likes of *it is raining heavily now* and *John is at a conference in Kiel*. These are not constant states, but the rain may stop at any point and John will leave Kiel eventually. Due to the temporary nature of these events, we cannot base any statement on general knowledge, but concrete (external) evidence is needed. The same applies to states-of-affairs that have not happened yet. However, also these states-of-affairs may become general knowledge for us after they have happened, and for example *it rained heavily on August 17, 2017* may be general knowledge for us, for example, because we will always remember this, since the rain ruined our garden party.

Even though the basic semantics of events contributes to the likelihood of something being general knowledge for us, it is important to bear in mind that the lack of external evidence is the key feature here. As stated above, the same piece of information may be based, e. g. on visual evidence, or it may be general knowledge for us. For example, the statement *Lisa is a vegetarian*

may be based on inference (at a barbecue Lisa eats only salad and tofu), or Lisa may have told us this herself. A different example is provided by, e.g. *the children are playing outside*. The differences in the information sources are very relevant, for example, in that only a statement based on recent visual evidence can be an answer to the question *where are the children?*. Whether these pieces of information are viewed as general knowledge or not is due to whether we refer to them via internal or external evidence (contested knowledge). In this respect, general knowledge resembles any other information source, since basically all our statements can be based on any kind of information source (our own emotions being a possible exception). However, general knowledge differs from other information sources in that the difference between general knowledge and any other information source is between internal and external evidence, while with, e.g. visual and hearsay evidence, the evidence is external in both cases. These differences are formally manifest, e.g. in Purik Tibetan:

(3) Purik Tibetan

- a. *zamb-e-r-i-ka pulispa jon-s duk/jot*
 bridge-GEN-DEM-GEN-LOC police.man come-PAST EX.T/EX.F
 ‘(I saw / I know that) there are policemen by the bridge.’
- b. *k^ho-a zbri-a rgjala duk/jot*
 (s)he-DAT write-INF good EX.T/EX.F
 ‘(I saw / I know that) (s)he’s good at writing.’
 (Zemp 2017: 269)

In both examples above, *duk* refers to contested knowledge, something the speaker has just learnt. For example, in (3a), the speaker has just seen some policemen near the bridge. *Jot*, in turn, presents the given piece of information as general knowledge. For example, in (3a), the speaker knows that policemen are usually standing by the bridge and s/he does not need any kind of other evidence for his/her claim (see Yliniemi 2017 for similar examples in Denjongke).

2.4 Neutralization/irrelevance of information source

Neutralization of the information source means basically that the original source of information has no bearing on the reliability of an utterance. This may have at least three different reasons. First, neutralization means that potential differences in the original information source are not relevant; speaker A may have learnt a piece of general knowledge by hearsay, while speaker B has visual

evidence for it. However, when the given information has genuinely become general knowledge for both, these differences are no longer relevant, and both speakers have absolute subjective certainty of the given piece of information. In this respect, general knowledge differs drastically from other information sources. For example, the claim *John is at home* is more reliable if we have visual evidence for this than if we are making a claim based on hearsay evidence. Neutralization follows very naturally here, because we may only have one kind of evidence (i. e. our own internal information) for general knowledge, while the nature of evidence for other statements varies.

Second, the speaker may have many different types of evidence (e. g. both hearsay and visual evidence) for general knowledge, and s/he cannot rank them according to their relevance. For example, in Wutun (see (6)), this often determines the use of factual evidentials (coding general knowledge) that are used whenever the original information source is not important, or the speaker is unable to specify it. Moreover, intersubjectivity plays a role in Wutun, and factual evidentials are more common whenever the speaker expects a certain piece of information to be general knowledge to the hearer as well (Erika Sandman, p.c.).

Third, a given piece of information may be a part of the common knowledge of the speaker's cultural environment, and no-one can state the original information source anymore. What matters at the present moment is the speaker's subjective absolute certainty, its exact source plays no role. There is thus no urgent need to specify the information source for general knowledge of this type, whereas this may be highly relevant for other information sources.

Even though the original information source has typically been neutralized for general knowledge in the ways discussed above, it is important to note that the original source may also affect the coding of general knowledge in some languages. For example, in Denjongke, the use of copulas varies according to whether a piece of general knowledge has become a part of the speaker's established world view via hearsay (something that everyone knows) or via personal experience; the copula *jəʔ* is used if the claim is based on personal experience, while *jəbbeʔ* appears if the speaker bases his/her claim on general knowledge heard from someone else (Yliniemi 2017).

2.5 Interim summary

In this section, general knowledge has been defined as pieces of information that have become the speaker's internal information, which renders the given

information certain and makes external evidence unnecessary. However, even though general knowledge can be given a seemingly homogeneous definition, it is important to note that it does not constitute an invariable evidence type, but we can distinguish between at least two different types of general knowledge. The first type, which probably constitutes the prototype of general knowledge, is presented by generally known truths, such as scientific, historical and geographical facts. The second type of general knowledge is illustrated by states-of-affairs that have become parts of our established world view based on our personal experiences of the world. Typical examples include *John is a linguist* and *Lisa supports the Pittsburgh Penguins*. As such, both types equally well represent the notion of general knowledge, but they also display differences.

First, the two types of general knowledge differ from each other in what kind of information we typically have, or can have, for them. We typically (yet not always) learn scientific facts in school or from other reliable sources, while we usually have personal experience for the second type. Moreover, the nature of evidence for the second type of general knowledge may vary, and we may claim that John has two children based on any type of evidence. On the other hand, we usually do not base our claims on scientific facts on, e. g. inference or assumption, but we learn them from an authority.¹ Consequently, we less often question scientific facts and they readily become general knowledge for us. On the other hand, a state-of-affairs becomes general knowledge of the second type easier, if we have reliable evidence for it. Differences in the original information source thus contribute to how easily something becomes general knowledge especially as regards the second type.

Second, the two types of general knowledge differ from each other in that contradicting evidence affects the truth value of the second type rather easily, while scientific facts are more stable in nature. For example, if we learn (from a reliable source) that John and Lisa have moved to Dunedin, the statement *John and Lisa live in Canberra* is no longer true, and we need to change our world view accordingly. On the other hand, if someone tells us that Paimio is the capital of Finland, or we see an anteater eating something else than ants, *Helsinki is the capital of Finland* and *anteater eats ants* nevertheless remain parts of our general knowledge. Moreover, we may presume generally known scientific facts to be known by the hearer as well, while with the second type of general knowledge this is not necessarily the case (see also Section 4.2 for a discussion of this difference).

¹ Scientists may naturally, at least at earlier stages of research, base their claims on inference.

The semantic differences between the two types of general knowledge discussed in this section are also linguistically relevant. First of all, scientific facts are most naturally referred to in the present tense and indicative mood. This also applies to the second type of general knowledge whenever a given state-of-affairs holds, but their coding is not as constant, since the states-of-affairs they refer to may change, and thus past tense is possible with them (*2 + 2 was 4* sounds odd, while *John and Lisa were married* does not). Second, the types of general knowledge differ in their coding as general knowledge. For example, in Wutun factual evidential occurs most often with the second type, e. g. to refer to states-of-affairs the speaker has learnt based on their frequent occurrence (Erika Sandman, p.c.). This distribution may seem somewhat unexpected at first, because the prototype of general knowledge is left unmarked. However, this becomes more natural if we consider the semantics of scientific facts. We know without any explicit coding that *2 + 2 is four* is a fact, which makes it unnecessary to underline this linguistically. In a similar vein, we may dispense with case marking if the semantic role assignment of the participants is clear, as in 'Lisa saw the rock'. Second, the differences between contested and uncontested evidence may be communicatively relevant, which may render it important to highlight the factual nature of our statement linguistically. This may be necessary also for other reasons, e. g. for underlining the general validity of the given piece of information. These differences do not exist for the first type of general knowledge.

In addition to the two discussed types of general knowledge, folklore, or traditional stories of one's own people, can also be seen as a type of general knowledge, because folklore is a part of the history of one's own people. Folklore resembles the first type of general knowledge in that we are usually not able to specify its original source, but there are also clear differences. One of the most important differences is represented by the story-like nature of folklore. Folklore may be viewed as stories of one's own people, which has clear consequences for their nature. From this it also follows that it is basically impossible to have any (dis)confirming evidence for it, which is in principle always available for the second type of general knowledge, and for a part of the first type. These differences are also linguistically manifest, and the coding of folklore differs drastically from the other types of general knowledge; general knowledge is consistently coded by direct evidentials, whereas there is massive variation for folklore, and indirect evidentials are far more frequent with folklore (see Kittilä in prep.). We may thus say that folklore is coded based on its origin (hearsay) rather than its reliability, which is a clear contrast to general knowledge. Due to the above mentioned semantic and formal differences, folklore is not discussed further in this paper.

3 A formal-functional typology of general knowledge coding

3.1 Preliminaries

In this section, a formal-functional typology of general knowledge coding is proposed based on the nature of the element used for this. Five types will be distinguished. In the typology, visual and direct evidentials have been lumped together because of their semantic similarity. Either label is used depending on the language and the terminological choices by the author, but usually the given marker comprises both direct and visual evidence (regardless of the label), which makes an explicit distinction unnecessary. In this section, the focus lies on the typology, the rationale behind the types will be discussed in Section 4.

A few methodological notes are in order. First, the discussion below only considers the coding of the two first types of general knowledge. This follows, because the status of the third type as general knowledge is questionable, as noted in Section 2.5. Considering the third type could yield a distorted picture of general knowledge coding. Potential differences in the coding of the two first types will not be discussed, because the goal is to illustrate the formal means languages employ for coding general knowledge. Moreover, differences between different types of general knowledge are usually not discussed in the consulted sources. Second, general knowledge coding is related to other linguistic elements as well, and it would thus be possible to discuss the formal means also, e.g. from the perspective of modality (indicative mood usually codes general knowledge). However, in this paper the focus lies solely on evidentiality, and any other related feature is not taken into account. It is also important to note that the proposed types differ in size, but this is considered secondary, because the goal is to illustrate the possible mechanisms. Finally, the typology is based only on markers of evidentiality that can somehow be considered grammatical. Evidential particles are (occasionally) considered, but purely lexical items, such as *definitely* and *without a doubt* fall outside the scope of this paper. Moreover, I will not discuss assertives, even though they can in many cases be used to underline the factual nature of a piece of information. This follows, because assertives are not necessary for referring to general knowledge, even though they may appear with it.

3.2 Type 1: Dedicated general knowledge marker

The first (and the least controversial) type is illustrated by languages that have a dedicated grammatical marker for the expression of general knowledge. Consider:

- (4) Mamainde
- a. $wa^3kon^3-\emptyset-na^2he^3-la^2$
work-3SG-VIS.PAST-PERF
'He worked (yesterday; I saw him)'
 - b. $wa^3kon^3-\emptyset-nü^2he^3-la^2$
work-3SG-INFR.PAST-PERF
'He worked (yesterday; I inferred this based on visual evidence)'
 - c. $ti^3ka^3l-a^2 \quad kai^3l-a \quad yain-\emptyset-te^2ju^2he^3-la^2$
anteater-DEF ant-DEF eat-3SG-GEN.KNOWLEDGE.EV-PERF
'The anteater habitually eats ants (I know this as general knowledge)'
 - d. $wa^3kon^3-\emptyset-ta^1hxai^2he^3-la^2$
work-3SG-REP.PAST-PERF
'He worked (I was told)'
(Kroeker 2001: 62–65, as cited in Aikhenvald 2004: 56)
- (5) Central Pomo
- a. $\check{c}^h\acute{e} \text{ mul}=?ma$
rain fall=GEN.KNOWLEDGE
'It rained (that's an established fact)'
 - b. $\check{c}^h\acute{e} \text{ mul}=ya$
rain fall=VIS
'It rained (I know because I was there and saw it)'
 - c. $\check{c}^h\acute{e} \text{ mul}=?do$
rain fall=REP
'It rained (I was told)'
 - d. $\check{c}^h\acute{e} \text{ mul}=nme$
rain fall=AUD
'It rained (I heard the drops on the roof)'
 - e. $\check{c}^h\acute{e} \text{ mul}=?ka$
rain fall=INFR
'It must have rained (everything is wet)'
(Mithun 1999: 181, glosses my own)

(6) Wutun

a. *ngu huan he-di-yek*

1SG food drink-PROGR-EGO

‘I am eating.’

b. *gu huan he-di-li*

3SG food drink-PROGR-SEN.INF

‘S/he is eating.’

c. *gu selang-ha lhojjhong-ghi-de re*

3SG Xining-FOC study-go.COMPL-ASS FACT

‘S/he will go to Xining for study (as we all know).’

(Sandman 2016: 207, Erika Sandman, p.c.)

Mamainde, Central Pomo and Wutun are all languages with an obligatory, grammaticalized evidentiality system. Regardless of the number of formally distinct evidentiality categories (which varies from three to five, as can be seen above), the languages above have a dedicated marker for general knowledge (labelled factual evidential in Sandman 2016). Similar cases are also attested in Mosuo (Lidz 2007: 60), Darma (Willis 2007: 96–97), many Tibetan languages (see, e. g. Zemp 2017: 262 for Purik Tibetan), and Japhug (Jacques 2015). Japhug presents an interesting case, because the general knowledge marker may also be used for coding events that are very likely to take place in the future.

A somewhat different example is provided by Kashaya Pomo illustrated in (7):

(7) Kashaya Pomo

a. *qowaq-wa*

pack-FACT

‘(I see) he is packing.’

b. *qowaq-ya*

pack-VIS

‘(I just saw) he packed.’

c. *si^hta=yachma cahno-w.*

bird=PL:SUBJ sound-FACT

‘Birds sing’ (general statement about birds) or ‘(I see/saw) birds are/were singing.’ (imperfective action)

(Oswalt 1986: 36, 244, as cited in de Haan 2001)

As the examples in (7a) and (7c) show, Kashaya Pomo has a factual evidential (Oswalt’s label) that can be used for referring to general knowledge, as shown

in (7c). What distinguishes Kashaya Pomo from the other languages discussed in this subsection is that factual evidentials can also code visual evidence (in which case they resemble Type 3 languages), which is manifested in the two readings given for (7c), and the reading of (7a). Moreover, the visual evidential ((7b)) cannot denote a generally known fact, but it appears with “perfective or completed actions which the speaker knows to be true because he saw them happen.” (Oswalt 1961: 244). Visual and factual evidential also display aspectual differences; factual evidential appears with imperfective and visual with perfective actions (Oswalt 1986: 36).

3.3 Type 2: Ego-evidential

The second type is illustrated by languages in which the evidential marker usually used for deliberate actions by a first person referent (ego-evidential)² is used for coding general knowledge as well. An illustrative example is provided in (8):

- (8) Oksapmin
- a. *nox [...]* *əbop dap=si dum-m sxa-sux*
 1S rope long=with tie-SEQ look.after-HAB.PER.FP.SG
 ‘I used to tie him up with rope and look after him.’
- b. *aw-xenil ixile dik j=olxol*
 grandparent.1POSS-PL 3P.POSS time DEM.DST=3SG.M.REFL
nuxul kukumi jox moxe-sxe
 1PL.EX bride.price DEF buy-HAB.PER.FP.PL
 ‘In the elders’ time, we used to pay bride price.’
 (Loughnane 2009: 254–257)

In (8a), the personal-factual marker *-sux* codes an event the speaker has been personally involved in and that s/he has initiated with intent. In (8b), an allomorph of the same marker is used for referring to general knowledge known to the people of the community. Functionally, general knowledge coding of Oksapmin is close to the languages of the first type in that both have a dedicated marker for general knowledge. However, in Oksapmin, the given marker is not confined to general knowledge coding, but it also covers purposefully instigated actions by a first person referent, i. e. it is not a pure factual/general knowledge evidential. It is also noteworthy that in Wutun (see (6)), and many Tibetan languages, these two

² The term used by Loughnane is personal-factual evidential, but I have replaced it with ego-evidential for coherence in terminology.

categories do not overlap, but ego-evidential and factual evidential are formally distinct. Type 2 seems to be rather rare across languages, even though similar cases are attested (the references cited as in Loughnane 2009: 251–253) also in Foe (Rule 1977: 74) and Fasu (Loeweke and May 1980).

A somewhat different manifestation of Type 2 is given in (9) (the language also has an inferential and a hearsay evidential):

(9) nDrapa

- a. $\bar{w}ore \bar{w}o \acute{ng}angu \bar{c}e \acute{cu}=q-\epsilon.$
 every.year tax give need=IPF-DISJ
 ‘We need to pay our taxes every year.’
- b. $\acute{j}en\lambda \bar{\eta}oro \bar{h}t\epsilon wu \bar{g}\lambda-?d\acute{j}-a.$
 yesterday 3SG PN DIR-arrive-[a]
 ‘He arrived at Tau yesterday (I saw him arrive).’
 (Shirai 2007: 133, 135)

In nDrapa, general knowledge is formally zero coded, as shown in (9a), while visual information is explicitly marked, as evidenced by the direct evidential in (9b). In principle, zero coding of general knowledge would render (9) an example of Type 4 (see below), but (9) is viewed as an instance of Type 2, because ego-evidential is zero coded in nDrapa, and the only difference between Oksapmin and nDrapa lies in the zero vs. explicit coding of ego-evidentials.

In this subsection, languages, where ego-evidentials code general knowledge have been discussed. Ego-evidentials are not always seen as a genuine information source, and they are, e. g. not considered in Aikhenvald’s (2004) seminal study. They have also been viewed as a part of person marking systems (see, e. g. Hale 1980). However, in this paper, ego-evidentials are seen as an information source in their own right, primarily for four reasons. First, they do constitute an obligatorily coded information source type in languages such as Wutun, and many Tibetan languages, similarly to, e. g. visual and hearsay evidence in other languages. Second, they can appear with other persons as well (with somewhat different meanings), which means that they do not constitute a mere person marking device. Third, in languages with grammaticalized ego-evidentials, ego-evidentials are in contrast with other information sources, and we can make statements based on our own volitional involvement in an event in the same way as our claims can be based on, e. g. hearsay evidence. Finally, ego-evidence resembles general knowledge in that in both cases the evidence is speaker-internal, even though in other respects these information sources display clear differences.

3.4 Type 3: Visual/direct evidential

The third type is illustrated by languages where general knowledge is coded by the visual/direct evidential marker of the given language. Examples are found in (10) and (11):

(10) Tucano

bũhĩpũu opâ-siti dî-bĩ
sun anaph.CL:ROUND be-PRES.VIS.3SG.M

‘The Sun is round’

(Ramirez 1997, vol. I: 127, as cited in Aikhenvald 2004: 170)

(11) Cuzco Quechua

Yunka-pi-n k'usillu-kuna-qa ka-n
rainforest-LOC-DIR monkey-PL-TOP be-3P

‘In the rainforest, there are monkeys’

(Faller 2002a: 20)

Despite the semantic differences between the denoted states-of-affairs, a visual/direct evidential is used for their coding. The employed markers are despite their obvious similarities, also somewhat different in nature. In Tucano, the marker of visual evidence also codes tense simultaneously, while in Cuzco Quechua, the given marker is a pure evidential. The use of visual/direct evidentials for general knowledge is rather common in languages and cases similar to those in (10)–(11) are also attested in, e. g. Tsafiki (Aikhenvald 2004: 172, information from Connie Dickinson), Tuyuca (Barnes 1984: 259), Shipibo-Konibo (Valenzuela 2003: 35), and Wanano (Stenzel 2008: 416).

3.5 Type 4: Zero marking

The fourth type is illustrated by languages in which general knowledge is coded by zero. Zero coding means that general knowledge is referred to by constructions without any kind of evidential marker regardless of whether the given language has some form of grammaticalized evidentiality or not. In addition, the general factual nature can be stressed by assertives or expressions such as *I know for a fact that ...* or *everybody knows that ...*. However, as noted also above, these are not considered here, because they are not necessary, but zero coding suffices. In principle, languages such as nDrapa (see also LaPolla 2003: 198–199 for Qiang) could also be classified as zero marking languages, but here the

notion is confined to languages where evidentiality is not in any way obligatorily coded, and optional evidentials do not appear with general knowledge. In languages like nDrapa and Qiang, zero marking is a part of the evidentiality paradigm with a clear function. In languages discussed in this section, in turn, zero is not directly related to any kind of evidentiality value, but the exact reading of zero is determined contextually. Two different examples of Type 4 are given in (12) and (13):

(12) Finnish

- a. *Maa kiertää Aurinko-a*
 Earth orbit.3SG.PRES sun-PART
 ‘The Earth orbits the Sun’
- b. *kaksi plus kaksi on neljä*
 two plus two COP.3SG.PRES four
 ‘Two plus two is four’
 (personal knowledge)

(13) Korean

- a. *welington-i nyucilaynd-ui suto-ta*
 Wellington-NOM NZ-GEN capital-IND
 ‘Wellington is the capital of New Zealand’
- b. *i tehaki i-un sa-ta*
 two adding two-TOP four-IND
 ‘Two plus two is four’
 (examples courtesy of Jae Jung Song)

Finnish and Korean differ in their evidentiality expression. In Finnish, evidentiality is not a grammaticalized category, but the information source is referred to by evidential particles (e. g. inferential and hearsay) or lexical verbs (‘see’, ‘hear’ etc.). Korean, in turn, displays grammaticalized evidentiality (verb affixes for direct, inferential and hearsay evidence), but, in contrast to the languages discussed above, their use is not obligatory (Jae Jung Song, p.c.). Despite these differences, general knowledge is formally zero coded in both languages; the Finnish evidential particles and the Korean evidential affixes do not appear with general knowledge. It should be noted that in Finnish and Korean (maybe along with many other languages of this type), zero marking is not directly and/or exclusively related to general knowledge. For example, in Finnish, zero also codes ego-evidence, visual evidence and most other instances of highly reliable information regardless of its exact nature. Zero coding can be considered the default, and it occurs if there is no need to specify that we are not dealing with

direct information. Moreover, it should be noted that optional markers of evidentiality are by no means excluded with general knowledge in Finnish and Korean. However, their appearance has obvious semantic consequences. For example, in Finnish the use of hearsay evidential creates the impression that the speaker has just learnt the given piece of information. Type 4 is common crosslinguistically, because basically all languages where evidentiality is an optional category represent Type 4.

3.6 Variable type

Finally, there are languages where the coding of general knowledge varies according to, e. g. the nature, or the origin of general knowledge. Two examples are illustrated in (14) and (15):

(14) Purik Tibetan

- a. *ŋa-a pene maŋmo jot*
 I-DAT money a.lot EX.F
 ‘(I saw / I know that) I have a lot of money.’
- b. *noskar t^haq-pa-na mar biŋ-et*
 rape grind-INF-CND oil come.out-CRT
 ‘When you grind rape(seed), oil comes out.’
 (Zemp 2017: 268, 281)

(15) Denjongke

- a. *bil geits=lo ny: kɛ:p jɛb-bɛ?*
 Bill Gates=LOC money a.lot COP.EXT.NMLZ-COP.EQU.NE
 ‘Bill Gates has a lot of money (as is generally known)’
- b. *k^hui=ki baik=di lɛpti mɔla? jɔ?*
 his=GEN bike=DEMPH very fast COP.EXT.PERS
 ‘His motorbike is very fast’
 (Yliniemi 2017)

In Purik Tibetan, the variation is between an existential copula and an affix, while in Denjongke, the variation concerns two different copulas. In Purik, the (epistemically neutral factual) existential copula *jot* contrasts with the testimonial existential copula *duk* used whenever the speaker is basing his/her claim on recent evidence. The prospective certain affix – *et*, in turn, is also used for referring to events that will certainly take place in the near future, i. e. the speaker has (subjective) certainty of both these event types. In Denjongke, the variation is between neutral and

personal copulas (terms from Yliniemi 2017). The neutral copula *jèb-be?* is used when the speaker is making a general statement without any reference to concrete evidence. This copula is thus a rather genuine general knowledge marker. The personal *jò?*, for its part, refers to the speaker's already existing knowledge, and it can be used only when it is possible to acquire personal information for the claim over time (it is therefore not possible with historical facts). General knowledge is often based on the speaker's previous experiences, which makes *jò* a general knowledge marker as well.

4 Rationale

4.1 Preliminaries

In this section, the rationale behind the attested language types, and the status of general knowledge as an information source will be scrutinized in more detail. Also the relevance of the discussion to our general understanding of evidentiality will be examined briefly. Before proceeding, it is important to note that, as noted above as well, the language types proposed in the previous section seem to differ enormously in size, which could naturally be taken as evidence for a better motivation of the frequent types. However, the goal here is to argue for the existence of all language types on a general level, whereby their frequencies are not central. Moreover, for example ego-evidentials are rather rare across languages, which naturally makes Type 2 less frequent, even though it seems that ego-evidentials rather typically code general knowledge if they exist in a language.

To begin with, the language types can be divided into two according to whether they stress the differences, or the common features general knowledge have with other information sources (Type 5 can be motivated in different ways). The occurrence of Types 1 and 4 can be accounted for by the peculiar nature of general knowledge; Type 1 has developed a specific marker for it, while languages of Type 4 leave it unmarked for evidentiality. The latter may follow, e. g. because general knowledge does not have any kind of determinable information source apart from the speaker's internal knowledge, which makes evidentials incompatible with it. Languages of Types 2 and 3, in turn, stress the similarities of general knowledge and other information sources. First, ego-evidentials and general knowledge are both the speaker's own internal information, which accounts for Type 2, and second, visual evidence and general knowledge are both highly reliable information, which explains the occurrence of Type 3.

The organization of this section is as follows. In 4.2, the close relation between general knowledge and reliability of evidence will be discussed, which will be followed by a discussion of general knowledge and (in)direct evidence in Section 4.3. In Section 4.4, I will discuss the status of general knowledge as an independent information source in more detail, while Section 4.5 discusses the theoretical implications this study has.

4.2 General knowledge and reliability of evidence

As noted above, general knowledge constitutes reliable information that the speaker believes to be true. The high degree of reliability is based on our previous experiences; we have deduced that certain states-of-affairs hold constantly, or we have learnt things about the world from books, authorities, or other (reliable) secondhand sources. General knowledge shares the high degree of reliability with visual evidence and ego-evidence. It is very hard to deny general knowledge/facts (especially of Type 1), or something we have done ourselves intentionally or something we have witnessed visually. The high degree of reliability also means that the speaker can take personal responsibility for claims based on these evidence types (see also Aikhenvald 2004: 189). All other kinds of evidence are somehow less reliable, and the degree of speaker responsibility is thus lower. For example, even though direct sensory evidence is usually necessary for inference, our inference may be wrong, which makes statements based on inference (and assumption), less reliable. Finally, in the case of hearsay evidence, we completely rely on information from others, which excludes personal responsibility for these statements. Consequently, markers of less direct evidence do not appear with general knowledge; the conveyed message would be misleading as regards its reliability.

Even though general knowledge shares reliability and high degree of speaker responsibility with ego-evidence and visual evidence, the motivation of reliability varies drastically between these evidence types. Statements based on ego- or visual evidence are based on concrete evidence. Moreover, the nature of the evidence is known to the speaker, and s/he can specify it. For general knowledge, in turn, the only type of evidence available is the speaker's own internal evidence and his/her view of the world. Consequently, the identical coding of general knowledge and ego-/visual evidence does not follow from the nature of evidence *per se*. It is interesting to note that in (8), where an ego-evidential is used for coding general knowledge, the statement involves a first person agent, i. e. the given piece of information concerns the speaker directly. This lends more support to the personal nature of general knowledge, and very well explains why ego-evidentials are used for coding general knowledge regardless of its origin.

Even though ego-evidence, visual evidence and general knowledge all constitute reliable evidence, differences may become evident if we have multiple types of contradicting evidence available at the same time. For example, if we have believed that John and Lisa have two children, but John himself tells us that they actually have three, we will (if we act rationally) change our world view accordingly. In other words, directly observable and presently available evidence outranks the less concrete evidence in this case. This is rather directly related to the subjective nature of general knowledge, and also aids us in explaining why subjective certainty may override non-linguistic reality and a non-fact may become general knowledge for us. Moreover, the status of a piece of information as general knowledge is not necessarily constant. We may forget things that used to be general knowledge for us, which makes external evidence necessary. On the other hand, visual evidence is in this sense more objective and harder to deny, which may result in us ranking it higher in case more types of evidence are available.

The two types of general knowledge discussed briefly above display clear differences in how other types of (highly reliable) evidence affects their truth value. Since some scientific facts, especially generally known and constant mathematic facts are very hard to deny, it is also difficult to have reliable contradicting evidence against them. We are, for example, unlikely to change our view on the fact $2 + 2$ is 4, even though someone would make a case for $2 + 2$ is 13. However, as scientists we are familiar with the fact that scientific research changes our views of our field, and if provided with reliable evidence, we are usually willing to change our thinking. For example, if someone is able to convince us that there is a language that only has nasal vowels, the universal *all languages with nasal vowels also have oral vowels* no longer is general knowledge for us. The second type of general knowledge is, as noted above, less constant, which has the consequence that it is also easier to have reliable contradicting evidence for them. For example, if *Bobby and Mary live in Boston* is general knowledge for us, and we learn from a reliable source that they have moved to Anchorage, the new piece of information replaces the old piece of information as our general knowledge. This can be seen as an instance of best possible evidence (Faller 2002b); our world view is based on the most reliable (and recent) evidence available.

4.3 General knowledge and direct/indirect evidentiality

Even though the coding of general knowledge could in principle be based on both direct and indirect evidence (it has features in common with both types), it comes as no surprise that (all?) languages treat general knowledge as direct

evidence formally. Most importantly, as noted in the previous section, general knowledge and direct evidence both present very reliable types of evidence, which very well accounts for the use of direct evidentials. The coding is based on the current information status of the given piece of information, i. e. it is relevant to convey the message that the denoted information is reliable. The directness of information is motivated very differently for general knowledge and, e. g. visual evidence, but in both cases the speaker has direct access to the information. Hearsay, inferential and assumptive evidentials, for their part, always constitute indirect information, and the use of indirect evidentials with general knowledge would lead to a wrong interpretation (see, e. g. Johanson 2003: 283 for a similar note on Turkic languages). Hearsay evidence can be highly reliable if it comes from a high authority, such as the police, but this is not manifested in its coding (cf. Plungian 2010: 37). Also the less concrete nature of general knowledge is irrelevant to its coding. Finally, miratives are also excluded from general knowledge coding, because mirativity is related to the newness of information, while general knowledge is by definition old information for the speaker.

Even though general knowledge is typically coded by direct evidentials, it is in order to note that indirect evidentials are by no means excluded in constructions describing general knowledge, but their use is related to functional differences, whose nature depends on the type of general knowledge. With the first type of general knowledge, the use of indirect evidentials usually stresses the fact that the given piece of evidence has just become general knowledge for the speaker, or that the speaker has doubts about its truth value. This follows, since the truth value of the given utterance does not depend in any way on the speaker's own evaluation of the state-of-affairs in question. With the second type of general knowledge, the use of indirect evidentials deprives general knowledge of its general knowledge status (the requirements for general knowledge status are no longer fulfilled). In both cases, the result is thus a change away from genuine general knowledge, but the exact nature of this varies.

As regards the directness of information, it is interesting to note that in some languages with two or more direct evidentials (e. g. ego-evidential and a visual evidential), the most direct evidential is employed for coding general knowledge. For example, in Oksapmin, ego-evidentials are employed for coding general knowledge instead of the visual-sensory evidential (see (8)). Of these, the ego-evidential codes more direct information (Loughnane 2009: 254–255, see also Shirai 2007: 133 for nDrapa). This also seems functionally natural, since ego-evidentials are more compatible with general knowledge due to the internal nature of these evidence types. In Qiang, general knowledge receives the same formal treatment as highly reliable visual evidence, while less reliable visual information is coded differently (LaPolla 2003: 198). This distribution is intriguing

in light of the fact that in case the speaker has both general knowledge and other highly reliable information at his/her disposal, s/he probably views, e. g. visual information as more reliable than something s/he believed to be a fact (see above). This is also related to the fact that it is easier to take full responsibility for something we have witnessed recently and we can directly refer to.

4.4 The status of general knowledge as an independent information source

To begin with, we may note that only a handful of languages accord general knowledge dedicated formal coding, but the majority of languages treat general knowledge as direct evidence. In other words, we only have little formal evidence for stating that general knowledge should be viewed as an evidence type of its own. However, two things should be taken into account here. First, also auditive is viewed as an independent evidence type in those languages where it is clearly an evidential category (see, e. g. Jalava 2017). Second, the very consistent coding of general knowledge as direct evidence could be taken as evidence for functional consistence of the notion; even though languages do not accord general knowledge distinct formal coding, they nevertheless treat it formally in a similar way across languages.

Semantic evidence for arguing for the independent nature of general knowledge is considerably stronger. The dual nature of general knowledge is relevant for this; general knowledge is highly reliable, yet non-concrete information. In other words, we have subjective certainty of general knowledge even though we do not have any kind of concrete directly observable evidence when referring to it. With other information types, reliability of evidence correlates with concreteness, and concrete, directly observable evidence, outranks less concrete evidence in reliability. Moreover, general knowledge differs from other information sources also in that the speaker can, at least to some extent, choose what kind of evidence s/he uses for his/her statement. For example, we can use our internal evidence for the statement *Lisa has a PhD in linguistics*, or we can prove this by visiting Lisa's homepage (see also Yliniemi 2017 for a similar remark on the use of Denjongke copulas *du?* and *be?*). On the other hand, we cannot choose what kind of evidence we have for *John is now traveling to Montreal*.

Second, as noted above, the nature of evidence changes when something becomes general knowledge; information is no longer coded based on its original source. On the other hand, e. g. hearsay evidence is always coded as hearsay evidence as long as it remains mere hearsay evidence. The speaker's focus is on the factual nature of the given piece of information. S/he does not need to care

about the original information source, because it does not affect the reliability of his/her claim. The changes in the information status of general knowledge correspond at least partly to what Yliniemi (2017) has described as information source vs. type of knowledge; the information source may be the same for general knowledge and other pieces of information, but the type of knowledge changes.

The heterogeneous nature of general knowledge, i. e. the contradiction between reliability and concreteness, has the consequence that general knowledge is harder to classify, e. g. based on directness of evidence (as on the list in Section 1) than other evidence types. Basically, any kind of sensory perception is necessarily direct evidence; we have direct access to it, and the evidence must be concretely present. On the other hand, we cannot have visual evidence for future events, or something we have not witnessed personally. Hearsay evidence, in turn, is always indirect and non-concrete evidence, because it constitutes other people's information. General knowledge thus has features of both direct and indirect evidence; reliability makes it similar to direct evidence, its internal nature to ego-evidence, while its non-concrete (non-observable) nature rather renders it indirect evidence. It also resembles assumption in that we do not have any concrete (external) evidence for it. However, based on its formal treatment, we may, if so desired, classify general knowledge as direct evidence rather than indirect evidence. Put another way, reliability outranks non-concreteness as the determining factor here.

Finally, general knowledge constitutes a normal evidence type among others in that we can make statements based on our general knowledge similarly to any other evidence type. For example, the statement *John and Lisa have two children* can be based on our general knowledge of the world, or we may have visual evidence for this state-of-affairs. This does not underline the peculiar nature of general knowledge in any way, but rather stresses the status of general knowledge as an information source type among all others.

4.5 Theoretical implications for our general understanding of evidentiality

The discussion in this section also contributes to our understanding of evidentiality more generally. First, and perhaps foremost, the discussion has shown that general knowledge constitutes an evidence type of its own, even though the discussion has also shown that most languages code it by an already existing mechanism. Semantic evidence for the distinction is more obvious and convincing. What distinguishes general knowledge from other evidence types is that

general knowledge is harder “to put on a map”, i. e. general knowledge does not constitute a homogeneous evidence type.

Second, this paper contributes to our understanding of evidentiality systems and the classification of evidence types. As shown above, most languages treat general knowledge as direct evidence formally. This stresses the distinction between direct and indirect evidence, and in so doing between reliable and non-reliable evidence, and the relevance of this distinction to the nature of evidentiality systems. Plungian (2010) bases his distinction between directness and indirectness on one hand, and personal vs. non-personal nature of the evidence, on the other. Languages, where general knowledge is coded as direct evidence, many instances of direct evidence are lumped together, and these languages formally distinguish between direct and indirect evidence regardless of its nature. In general, languages seem to make more distinctions in the coding of indirect evidence, as shown by Aikhenvald (2004: Ch. 2). There are also languages, where the only explicit formal distinction is based on directness of evidence (like Cherokee and Yukaghir, discussed in Aikhenvald 2004: 26–27). This distinction is also clear in Finnish, where information source is left implicit for all instances of reliable/direct evidence including general knowledge. On the other hand, Finnish has evidential particles for different instances of less reliable evidence. We may also add that the differences between direct and indirect (and/or reliable and non-reliable) evidence are communicatively more relevant than the differences between, e. g. visual evidence and general knowledge.

The nature of evidentiality systems is also closely related to the emergence of evidentiality systems. Dedicated general knowledge/factual evidentials are, as illustrated in Section 3, rather uncommon across languages, and in my data there are no languages where this kind of evidential would be the sole evidential. This means that general knowledge evidentials do not constitute an early stage of development. The development rather starts with more obvious and relevant differences, and the first evidential to emerge is some kind of second-hand/hearsay evidential (e. g. Estonian only has a grammaticalized marker for hearsay evidence). As argued by Faller (2002b), speakers use the best possible evidence for their claims, which is in typical cases illustrated by direct/visual evidence, in many languages also including general knowledge. Hearsay evidence is maximally distinct from direct evidence, which makes it important to underline this difference explicitly. Similarly, languages with a dual and a plural first grammaticalize plural marking, since it is maximally different from the singular. On the other hand, general knowledge has many features in common with other instances of direct evidence, which makes explicit coding of this

difference communicatively redundant, and thus rather rare across languages. This also manifests their late development.

5 Conclusions

In this paper, the status of general knowledge as a source of information has been discussed. General knowledge was defined as pieces of information that the speaker has absolute subjective certainty of without having to resort to some external evidence, such as visual or hearsay evidence at the time when a sentence coding general knowledge is uttered. General knowledge has become an established part of the speaker's world view, which renders additional evidence redundant. The semantic nature of general knowledge speaks in favor of viewing it as an independent information source, but general knowledge also has features in common with other information sources.

In addition to discussing the semantics of general knowledge, a formal-functional typology of its coding was proposed. Languages were divided into five types according to the nature of the evidential marker used for referring to general knowledge; languages either have a dedicated marker for them, they use another (visual/direct or ego) evidential for coding them, or general knowledge is coded by zero (lack of evidentials). No languages were found in which inferential, assumptive or hearsay evidentials would be employed for this (this is attested for folklore, but this type of general knowledge was not taken into account in this study). The sample used for this study is, however, rather small, which means that the occurrence of languages where hearsay evidentials are used for the coding of general knowledge cannot be completely excluded. More detailed studies are needed to (dis)confirm this, as well as to shed more light on the actual frequencies of the five language types of the proposed typology. The five types differ from each other in whether they stress the similarities with or differences from other information sources; Types 1 and 4 stress the differences, while Types 2 and 3 are rather based the features general knowledge shares with other information source, most notably the reliability of information. In Type 5, this may vary according to the types involved in the variation.

The present study is only the first step towards a better understanding of general knowledge as an evidential category. Many aspects have only been touched upon, or not discussed at all in this brief overview. It is therefore my hope that other studies will follow on all possible aspects of general knowledge. First of all, it would be important to study the frequencies of the types in more detail; is it indeed the case that Type 1 is not very common across languages,

while Type 4 is attested frequently? Is Type 5 attested in the majority of languages, or at least in a higher number of them? Does the proposed typology exhaust the attested types, or is yet a further coding strategy possible? How often can zero be used for coding general knowledge in languages, where evidentiality is basically an obligatory category? What kind of functions do other than ego- or direct/visual evidentials acquire in languages where evidentiality is an obligatory category? All of these questions lie outside the scope of the present paper, but it is my sincere hope that other scholars working on evidentiality will discuss them in the future, both within and across languages.

Appendix: Abbreviations

ANAPH	Anaphora
ASS	Assertive
ASSUM	Assumed
AUD	Auditive
CL	Classifier
CND	Conditional
COMPL	Completive
COP	Copula
CRT	Certainitive
DAT	Dative case
DEF	Definite
DEM	Demonstrative
DEMPH	Deictic emphatic
DIR	Direct evidential
DISJ	Disjunct
DST	Distal
EGO	Ego-evidential
EQU	Equative
EV	Evidential
EX	Exclusive
EXT	Existential
EX.F	Factual existential copula
EX.T	Testimonial existential copula
FACT	Factual
FOC	Focus
FP	Far past
GEN	Genitive
HAB	Habitual
IND	Indicative
INF	Infinitive
INFR	Inferential

IPF	Imperfective
LOC	Locative
M	Masculine
NE	Neutral
NF	Non-feminine
NMLZ	Nominalization
NOM	Nominative
NONVIS	Non-visual
P	Person
PART	Partitive
PAST	Past tense
PER	Personal-factual evidential
PERF	Perfect
PERS	Personal
PL	Plural
PN	Personal name
POSS	Possessive
PRES	Present tense
PROGR	Progressive
REC.P	Recent past
REFL	Reflexive
REP	Reported
SEN.INF	Sensory-inferential
SEQ	Sequential
SG	Singular
SUBJ	Subject
TOP	Topic
VIS	Visual

References

- Aikhenvald, Alexandra Y. 2002. *Language contact in Amazonia*. Oxford: Oxford University Press.
- Aikhenvald, Alexandra Y. 2004. *Evidentiality*. Oxford: Oxford University Press.
- Aksu-Koç, Ayhan A. & Dan I. Slobin 1986. A psychological account of the development and use of evidentials in Turkish. In Wallace Chafe & Johanna Nichols (eds.), *Evidentiality: The linguistic coding of epistemology*, 159–167. Norwood, NJ: Ablex.
- Barnes, Janet 1984. Evidentials in the Tuyuca verb. *International Journal of American Linguistics* 50(2). 255–271.
- De Haan, Ferdinand. 2001. The place of inference within the evidential system. *International Journal of American Linguistics* 67(2). 193–219.
- Eberhard, David. M. 2012. The Mamaindê tense/evidentiality system. *Word Structure* 5(2). 129–164.
- Eunice, Loeweke & Jean May. 1980. General grammar of Fasu (Namo Me). *Workpapers in Papua New Guinea Languages* 27. 5–106.

- Faller, Martina. 2002a. *Semantics and pragmatics of evidentials in Cuzco Quechua*. Stanford, CA: Stanford University dissertation.
- Faller, Martina. 2002b. Remarks on evidential hierarchies. CSLI: *Proceedings of the SemFest*. 37–59.
- Floyd, Simeon, Elisabeth Norcliffe & Lila San Roque (eds.). 2018. *Egophoricity*. Amsterdam & Philadelphia: John Benjamins.
- Gawne, Lauren & Nathan Hill. 2017a. The contribution of Tibetan languages to the study of evidentiality. In Lauren Gawne & Nathan Hill (eds.), *Evidential systems of Tibetan languages*, 1–38. Berlin & Boston: De Gruyter Mouton.
- Gawne, Lauren & Nathan Hill (eds.). 2017b. *Evidential systems of Tibetan languages*. Berlin & Boston: De Gruyter Mouton.
- Hale, Austin. 1980. Person markers: Finite conjunct and disjunct forms in Newari. In Ronald L. Trail (ed.), *Papers in Southeast Asian linguistics*, 7, 95–106. Canberra: Pacific Linguistics.
- Jacques, Guillaume. 2015. Evidentiality in Japhug. https://www.academia.edu/14108663/Evidentiality_in_Japhug (accessed 18 February 2017).
- Jalava, Lotta. 2017. Grammaticalization of modality and evidentiality in Tundra Nenets. In Kees Hengeveld, Heiko Narrog & H. Hella Olbertz (eds.), *The grammaticalization of tense, aspect, modality and evidentiality: A functional perspective*, 133–162. Berlin & Boston: De Gruyter Mouton.
- Johanson, Lars. 2003. Evidentiality in Turkic. In Alexandra Y. Aikhenvald & R. M. W. Dixon (eds.), *Studies in evidentiality*, 273–290. Amsterdam & Philadelphia: John Benjamins.
- Kittilä, Seppo. In prep. Folklore as an evidential category.
- Kroeker, M. 2001. A descriptive grammar of Nambiquara. *International Journal of American Linguistics* 67. 1–87.
- LaPolla, Randy. 2003. *A grammar of Qiang*. Berlin & New York: Mouton de Gruyter.
- Lidz, Liberty A. 2007. Evidentiality in Yongnging Na (Mosuo). *Linguistics of the Tibeto-Burman Area* 30(2). 45–87.
- Loughnane, Robyn. 2009. *A grammar of Oksapmin*. Melbourne: University of Melbourne dissertation.
- Maslova, Elena. 2003. Evidentiality in Kolyma Yukaghir. In Alexandra Y. Aikhenvald & R. M. W. Dixon (eds.), *Studies in evidentiality*, 219–236. Amsterdam & Philadelphia: John Benjamins.
- Mithun, Marianne. 1999. *The languages of native North America*. Cambridge: Cambridge University Press.
- Oisel, Guillaume. 2013. *Morphosyntaxe et sémantique des auxiliaires et des connecteurs du Tibétain Littéraire*. Paris: Université Sorbonne Nouvelle-Paris 3 dissertation.
- Oswalt, Robert L. 1961. *A Kashaya grammar (Southwestern Pomo)*. Berkeley, CA: University of California Berkeley dissertation.
- Oswalt, Robert L. 1986. The evidential system of Kashaya. In Wallace Chafe & Johanna Nichols (eds.), *Evidentiality: The linguistic coding of epistemology*, 29–45. Norwood, NJ: Ablex.
- Plungian, Vladimir. 2010. Types of verbal evidentiality marking: an overview. In Gabriele Diewald & Elena Smirnova (eds.), *The Linguistic realization of evidentiality in European languages*, 15–58. Berlin & New York: De Gruyter Mouton.
- Ramirez, Henri. 1997. *A Fala Tukano dos Yepâ-masa*, 3 vols. Manaus: Inspeitoria Salesiana.
- Rule, W. M. 1977. *A comparative study of the Foe, Huli and Pole Languages of Papua New Guinea*. (Oceania Linguistic Monographs 20). Sydney: University of Sydney.
- Sandman, Erika. 2016. *A grammar of Wutun*. Helsinki: University of Helsinki dissertation.

- Shirai, Satoko. 2007. Evidentials and evidential-like categories in nDrapa. *Linguistics of the Tibeto-Burman Area* 30(2). 125–150.
- Stenzel, Kristine. 2008. Evidentials and clause modality in Wanano. *Studies in Language* 32(2). 405–445.
- Valenzuela, Pilar. 2003. Evidentiality in Shipibo-Konibo, with a comparative overview of the category in Panoan. In Alexandra Y. Aikhenvald & R. M. W. Dixon (eds.), *Studies in evidentiality*, 33–62. Amsterdam & Philadelphia: John Benjamins.
- Willett, Thomas. 1988. A cross-linguistic survey of the grammaticalization of evidentiality. *Studies in Language* 12(1). 51–97.
- Willis, Christina. 2007. Evidentiality in Darma (Tibeto-Burman). *Linguistics of the Tibeto-Burman Area* 30(2). 89–124.
- Yliniemi, Juha. 2017. Copulas in Denjongke or Sikkimese Bhutia. In Lauren Gawne & Nathan Hill (eds.), *Evidential systems of Tibetan languages*, 297–350. Berlin & Boston: De Gruyter Mouton.
- Zemp, Marius. 2017. Evidentiality in Purik Tibetan. In Lauren Gawne & Nathan Hill (eds.), *Evidential systems of Tibetan languages*, 261–296. Berlin & Boston: De Gruyter Mouton.