Consequences of evidential marking on the interpretation of subjective predicates: Experimental data from Korean[1]

Sarah Hye-yeon LEE — *University of Southern California*
Elsi KAISER — *University of Southern California*

Abstract. We report two psycholinguistic experiments on how two kinds of subjective predicates – predicates of personal taste (PPTs) and non-PPT multidimensional adjectives (MDs) – interact with evidential morphology in Korean. We build on theoretical discussion on PPTs and MDs, according to which the semantics of PPTs, but not MDs, makes reference to an individual's first-person experience. We test whether and how the perceived generalizability of opinions containing PPTs or MDs is modulated by evidentiality. We tested three evidential morphemes in Korean: the direct evidential *-te-*, the hearsay (reportative) evidential *-tay*, and the inferential evidential *-napo-*, along with a baseline condition with no evidential marking. The results suggest that (a) with PPTs, type of evidential marking has no effect on the perceived generalizability of opinions, but (b) with MDs, use of the inferential evidential lowers perceived generalizability. We attribute these patterns to the first-person direct experience requirement of PPTs (which we suggest renders opinions with PPTs harder to generalize, regardless of evidential marking), and to the decreased reliability of inferential evidence. Our work provides novel evidence for a link between evidentiality, subjectivity and the perceived generalizability of opinions.

Keywords: evidentiality, subjectivity, predicates of personal taste, multidimensional adjectives, Korean, generalizability, experimental semantics and pragmatics

1. Introduction

There exists an intuitive distinction between statements about objective facts (e.g. (1)) and those that express opinions (e.g. (2)).

(1)   It's raining in on USC's campus right now.
(2)   Frog legs are tasty.

The truth of statements like (1) can be objectively evaluated based on the state-of-affairs of the world. Consider, for example, (3). Here, one of the speakers must be in the wrong ('at fault'). Subjective opinions like (2), on the other hand, are not evaluated in exactly the same way. They give rise to *faultless disgreement* (e.g. Kölbel, 2004; Lasersohn, 2005, 2009; Stephenson, 2007a,b; Stojanovic, 2007; Pearson, 2013), exemplified in (4). Here, the speakers can disagree about the tastiness of frog legs without anyone being in the wrong. That is, when it comes to subjective matters, both speakers are entitled to their own opinions.

(3)   Speaker A: It's raining on USC's campus right now.
       Speaker B: No, it's not raining. It's sunny.

(4)   Speaker A: Frog legs are tasty.
       Speaker B: No, they're not tasty. They're disgusting.

Subjective predicates are analyzed by making reference to an individual's (or group's) subjective perspective or experience. Their subjective nature has been formalized in different ways in prior work, including accounts making use of a judge parameter (e.g. Lasersohn, 2005, 2009), truth-relativist accounts (e.g. MacFarlane, 2006), genericity-based accounts (e.g. Moltmann, 2010; Pearson, 2013), and outlook-based frameworks (e.g. Coppock, 2018). In the present paper, we do not aim to assess different theoretical accounts of subjective predicates, but we draw upon the general insight that the semantics of subjective predicates make reference to the perspective of an opinion-holder or 'judge.' (We use the term judge in an a-theoretic sense, without committing to formalizing it in a certain way).

Evidentiality is another dimension where an individual's perspective is relevant (e.g. Bylinina, Sudo, & McCready 2014). Evidentiality is a grammatical device that indicates the source of information (e.g. Chafe & Nichols, 1986; Willett, 1988; Aikhenvald, 2004, among many others), for example, whether the evidence holder has direct evidence, inferential evidence, or hearsay (reportative) evidence for the information. In some sense, evidentials are inherently subjective, as the evidence is possessed by a particular individual, usually the speaker, and thus their interpretation is typically analyzed as making reference to an 'evidence holder.'

Several scholars have recently explored the link between subjectivity and evidentiality (e.g. Nuyts, 2001; Korotkova, 2016; Anand & Korotkova 2018; Muñoz, 2020). We expand on this literature by testing how evidential markers in Korean guide the interpretation of subjective opinions, in particular *the perceived generalizability of opinions* (discussed more in Section 3.1). We tested two types of subjective adjectives: predicates of personal taste (PPTs) and non-PPT multidimensional adjectives (MDs). In what follows, we introduce some background and relevant theoretical discussion on the distinction between PPTs and MDs (Section 2) and on Korean evidentials (Section 3). We report findings from our psycholinguistic experiments in Section 4. We discuss conclusions and directions for future work in Section 5.

2. Two types of subjectivity: Predicates of personal taste and multidimensional adjectives

In this section, we review prior work on the two types of subjective predicates that we tested in our experiments: predicates of personal taste (PPTs) and non-PPT multidimensional adjectives (MDs). We first discuss claims that PPTs require the judge to have the relevant kind of *direct, first-person experience*, unlike MDs (Section 2.1). We then discuss the idea that the grounds for an opinion expressed with an MD can be linguistically communicated and justified to another person more easily than the intrinsically experiential opinions expressed with PPTs (Section 2.2).

2.1 The 'direct experience' requirement of PPTs

It has been argued that there are at least two subtypes within the broader class of subjective adjectives, namely predicates of personal taste (PPTs) and non-PPT multidimensional

adjectives (MDs, e.g. Sassoon, 2012; Bylinina, 2014; McNally & Stojanovic, 2017). In this section, we first review prior work on PPTs and then turn to MDs. The class of PPTs includes adjectives such as *tasty*, *fun*, *frightening*, and *irritating* (e.g. Lasersohn, 2005; Stephenson, 2007; Stojanovic, 2007; Anand, 2009; Pearson, 2013). With PPTs, it has been claimed that subjectivity is tied to direct, first-person experience (e.g. Stephenson, 2007a; 2007b; Anand, 2009; Pearson, 2013), which is why examples like (5) are perceived as infelicitous.

(5)   The cake is tasty, # but I have not tasted it.

Pearson (2013: 117) captures the first-person orientation of PPTs by stating that:

> 'In order to assert that *x* is P for some taste predicate P, one typically must have direct sensory experience of the relevant kind on the basis of which to judge whether *x* is P. For *tasty*, for example, I must have tasted the object I am talking about [...] If I have good reason to believe that shortbread is tasty, say because a reliable expert has told me so, I might say, *Apparently, shortbread is tasty*, but not, *Shortbread is tasty*'.

The relevance of direct, first-person experience has played an important role in theorizing about PPTs (e.g. Bylinina, 2014, Ninan, 2014). Intuitively, the phenomenon of faultless disagreement with PPTs (e.g. (4)) can be regarded as having its roots in divergent personal experiences: Two people's disagreement arises from their experiences being different. Indeed, Bylinina (2014) points out that the "presence or absence of an experiencer argument is orthogonal to judge-dependence" (Bylinina, 2014: 50) – in other words, judge-dependent expressions do not necessarily have to make reference to direct, first-person experience. She proposes a separate 'judge=experiencer' requirement, which states that "A direct statement about someone's internal state can be made only if the judge parameter is set to the same value as the experiencer of this internal state" (Bylinina, 2014: 58). According to Bylinina, this requirement applies to, but is not limited to, PPTs. She gives the semantics in (6) for *tasty*, where the second part states that the judge j and experiencer z refer to the same individual. Thus, according to Bylinina, adjectives like *tasty* have an experiencer requirement as part of their semantics, in contrast to other subjective adjectives, such as non-PPT multidimensional adjectives, e.g. *smart/healthy*, which she argues do not refer to experience in the same way.

(6) $[\![\text{tasty}]\!]^{c;w,t,\text{j}} =$
        (i) $\lambda z \lambda x.\ \exists s\ [\text{taste}(s)\ \&\ \text{Experiencer}(s, z)\ \&$
            $\text{Stimulus}(s, x)\ \&\ \text{TASTE}(s) \succ d_{st}\ \text{for } j \text{ at } t \text{ in } w];$
        (ii) JUDGE=EXPERIENCER: $j = z$

Along similar lines, Ninan (2014) discusses the relevance of direct, first-person experience for PPTs. Ninan observes that utterances of simple sentences containing PPTs usually trigger an *acquaintance inference*, which suggest that the speaker has relevant firsthand experience of the thing in question (see also Stephenson, 2007a; 2007b; Anand, 2009; Pearson, 2013; Klecha, 2014; Kennedy & Willer, 2016; Bylinina, 2017; Rudolph, 2019; Willer & Kennedy, 2021; cf. exocentric readings, Stephenson, 2007a; 2007b).

In recent work, McNally and Stojanovic (2017) also emphasize the importance of first-person experience in the case of PPTs. In line with Bylinina and Ninan, they note that PPTs involve

"the presence (or absence) of an experiencer, that is, a sentient individual who perceives the property in question" (McNally & Stojanovic, 2017: 24).

In sum, the information expressed by PPTs relies on the judge (typically the first-person speaker) having the relevant kind of direct experience that allows them to make the judgement/opinion being expressed. In this regard, PPTs differ from adjectives such as *gluten-free* or *healthy*. As will become clearer in the following sections, although adjectives like *healthy* are also subjective, a person can reach a judgement/have an opinion about whether something is healthy *without* having direct, first-person experience. For example, this could happen on the basis of hearing someone else's report or by inferring it from other types of evidence available to the judge.

To better understand the subjectivity of MD adjectives like *healthy*, let us take a closer look at their semantics. These adjectives belong to the second type of adjectives that we investigate, namely non-PPT multidimensional adjectives (MDs).[2] These are adjectives such as *healthy*, *intelligent*, and *useful* (e.g. Kamp, 1975; Klein, 1980; Bierwisch, 1989; Sassoon, 2013; Kennedy, 2013; McNally & Stojanovic, 2017). As their name suggests, in the case of MDs, multiple criteria can be used for ranking/ordering individuals with the property. Whether someone is intelligent, for example, can be evaluated along multiple dimensions such as math skills, language skills, survival skills, etc. Similarly, for evaluating whether someone is healthy, multiple criteria such as the state of her cardiovascular system, nervous system, immune system, etc. can be considered (see e.g. McNally & Stojanovic 2017 for discussion). Thus, with MDs, subjectivity stems from the fact that individuals can disagree about the weight of each dimension (McNally & Stojanovic, 2017, see also Sassoon, 2013; Kennedy, 2013; Bylinina, 2014).

Crucially for our purposes, it has been claimed that although MDs are subjective, they do not have the same kind of direct, first-person experience requirement as PPTs (e.g. Bylinina, 2014; McNally & Stojanovic, 2017, see also Kaiser & Herron Lee, 2017, 2018 for psycholinguistic evidence regarding the distinction between PPTs and MDs). For example, McNally and Stojanovic highlight the importance of distinguishing "predicates that entail a proper experiencer from those whose subjective element may be due to differences of opinion over where the threshold for ascribing a property lies or what the relative weights of the different criteria for ascribing it should be" (McNally & Stojanovic, 2017: 28).

Our hypotheses (Section 4) build on the intuition that the judge of a PPT needs to have direct, first-person experience, whereas the judge of MDs is not subject to the same requirement.

2.2 Multidimensional adjectives and indirect evidence

Let us now take a closer look at the idea that the judges of MDs are not subject to the same kind of first-person direct experience requirement as PPT judges are. Specifically, it seems that compared to the intrinsically experiential opinions expressed with PPTs, the grounds for an

---

[2] It is worth noting that some PPTs could perhaps be argued to be multidimensional. One may find a movie interesting in some aspects, but not in some other aspects (dimensions). However, this is not central for the aims of the present paper. Here, we compare PPTs to non-PPT multidimensional adjectives.

opinion expressed with an MD can be linguistically communicated and justified to another person more easily. According to Vardomskaya (2014), the key difference between PPTs and MDs is that the former, but not the latter, have dimensions that are '*communicable*'. In other words, the criteria a speaker has for holding an opinion expressed by an MD can be shared with others/communicated in a way that the reasons for a PPT-type opinion cannot be. This is because PPTs are based on first-person, internal experience. To determine whether a cake is tasty, you have to gain *direct, first-person experience* and taste it for yourself. In contrast, to determine whether someone is healthy or smart, you can rely on *second-hand reports and other types of indirect evidence*, without needing direct first-person experience. For example, at least in certain contexts, someone can judge another person to be smart based on information about the person's IQ, GRE, or SAT score, and someone can judge another person to be healthy based on reports of that person being able to run X meters in Y seconds (e.g. when assessing army recruits).[3]

The relatively easier communicability of MDs, along with the observation that the judge of MDs is not required to have first-person experience, suggest that even individuals who do not have direct experience concerning the thing in question can form MD opinions based on second-hand reports or other types of indirect evidence – unlike the situation with PPTs, which require direct first-person experience. This asymmetry between PPTs and MDs will become highly relevant once we consider how different kinds of evidentials impact their interpretation.

3. Generalizability of opinions and evidentiality

The experiments in the present paper investigate how evidential markers guide the perceived generalizability of sentences containing subjective adjectives (PPTs and MDs). In thinking about opinion generalizability, we take as our starting point the assumption that opinions differ in how prevalent they are. While some opinions are widespread and shared by many people, others are held by a smaller number of individuals. It is worth emphasizing that our experiments only test the *perceived* generalizability of opinions, *not* their actual prevalence. In the present work we do not test how many people actually hold a particular opinion (see Kaiser & Rudin, 2020 for work on that topic); rather, we assess how the presence of evidential markers and adjective types influences comprehenders' assessments of how generalizable they *think* an opinion is. We call this *perceived generalizability*.

Our focus on opinion generalizability stems from work by Moltmann (2010) and Pearson (2013), who claim that PPTs express not only first-person opinions but also convey information about 'people in general.' In what follows, we first review prior literature on the link between subjective adjectives and genericity, along with our own earlier work on the relation between opinion generalizability and subjective adjectives (Section 3.1). In Section 3.2, we suggest that the perceived generalizability of opinions can be modulated by differences in evidentials.

3.1 Generalizing opinions beyond the first-person speaker

---

[3] We put aside for now related questions concerning notions such as social consensus and social norms (see e.g. Kaiser & Rudin, 2020) and external authorities/definitions (see e.g. Vardomskaya, 2014 for discussion).

So far, we have focused on the importance of first-person experience with PPTs. However, Moltmann (2010) and Pearson (2013) claim that, in addition to expressing a first-person opinion, PPTs can also be interpreted as statements about 'people in general' (see also Bhatt & Izvorski, 1995; Collins, 2013; Snyder, 2013). In the words of Pearson (2013), PPTs "are used to make statements about whether something is tasty to people in general, based on first person experience" (Pearson, 2013: 121). Relatedly, Moltmann (2010) describes PPTs as expressing a generalization that is "to be true with respect to anyone of the relevant sort (including the interlocutors)" (Moltmann, 2010: 210). Moltmann proposes that PPTs involve a special kind of first-person oriented genericity, which has already been proposed for generic *one* and arbitrary PRO (Moltmann, 2006; 2010).

In our earlier experimental work on English (Kaiser and Lee, 2019; 2021), we tested whether the perceived generalizability of opinions expressed with PPTs and MDs depends on the explicitness of the opinion holder, the valence of the predicate, and the generic vs. episodic nature of the sentence. To do this, we developed an alien planet task. Participants were asked to imagine that they were visiting an alien planet and heard aliens talking to each other about things on the alien planet. For example, an alien might say to another alien "Churbits are tasty" or "That zirby was frightening" or "I thought that blicket was irritating." We used nonce words to eliminate any effects of participants' own opinions about real things (see e.g. Fazio et al., 1986). We operationalized perceived opinion generalizability as the number of other aliens that a participant thinks would share the alien's opinion. Participants indicated their estimation of how many other aliens, out of 100 randomly selected aliens on that planet, share the opinion of the alien who made the critical statement. In those studies, as in the present work, we tested the *perceived* generalizability of opinions, rather than their actual generalizability (which is in fact unknown in the alien context.) The results of this earlier work suggest that opinions expressed in matrix clauses (e.g. "That churbit was tasty") are perceived as more generalizable than ones embedded under *think/find* (e.g. "I thought/found that churbit was tasty"). We also found differences in the generalizability of positive vs. negative opinions, but (surprisingly) no stable differences between generic ("Churbits are tasty") vs. episodic ("That churbit was tasty") sentences in terms of the perceived generalizability of opinions.

Our earlier work shows that the alien task – which we also use in the studies reported here – yields interpretable data about the perceived generalizability of opinions. However, our earlier results do not speak to the relation between perceived opinion generalizability and type of evidence available to the speaker. The question of whether opinions are perceived as more or less generalizable depending on whether the speaker has direct or indirect evidence for their opinion is still open, as is the question of whether this interacts with adjective type (PPT/MD).

3.2 Evidential strength

To address the open questions identified above, we test whether the perceived generalizability of opinions is modulated by grammatical markers in a linguistic domain closely related to subjectivity: evidentiality. There is a pragmatic intuition that some evidentials are perceived to be "stronger" than others. This intuition is captured in *evidential hierarchies* (e.g. Willett, 1988; de Haan, 1998; Faller, 2002; Davis, Potts, & Speas, 2007), linear scales on which evidentials (or evidence types) are ordered according to the reliability (or a similar concept) of

the source of information. Thus, an evidential hierarchy essentially represents how strongly a speaker knows her claim. A common intuition underlying various evidential hierarchies is that direct evidence is stronger than indirect evidence, as in (7).

(7)   Direct evidence > Indirect evidence

Davis, Potts, and Speas (2007) model the intuition that these hierarchies are subject to contextual variation (see also e.g. Faller, 2002). For instance, although direct evidence is usually more reliable than hearsay evidence, there might exist contexts in which hearsay evidence is more reliable than direct evidence (e.g. if the hearsay source is highly trustworthy).

To the extent of our knowledge, prior work has not discussed the evidential hierarchy or evidential strength in the context of subjective opinions. We investigate whether evidentiality, or evidential strength broadly construed, modulates the perceived generalizability of subjective opinions expressed with PPTs and MD, by testing the interpretational consequences of different evidential markers in Korean.

3.3 Korean evidentials

Korean uses optional verbal suffixes to mark the source of information. Broadly speaking, Korean has three main types of evidentials: direct evidentials, inferential evidentials, and hearsay (or reportative) evidentials (Chung, 2005). The Korean direct evidential marker *-te-* (e.g. Yang, 1972; Seo, 1977; Kim, 1981; Song, 1998, 2002; Chung, 2006, 2007; Kwon, 2009; Lee, 2010; Lim, 2010) is analyzed as marking direct perceptual evidence (e.g. Lee, 2010). The morpheme *-te-* functions as a direct evidential when it appears in a present tensed sentence (Lee, 2013). (All of our experimental stimuli were in present tense). The sentence in (8), for example, can be uttered when the speaker has direct perceptual evidence of the rain (e.g. visual evidence).

(8)   pi-ka   o-te-la
       rain-NOM   come-DIR.EV-DECL
       'It rained (, I have direct evidence.)'

The Korean hearsay (reportative) evidential marker *-tay* marks information reported by a third-party speaker (e.g. Chung, 2009). We note that for Korean hearsay evidentials, the hearsay source is underspecified: it can be another individual, multiple individuals, or even social consensus (e.g. 'It is said that…'). For example, (9) is felicitous if the speaker heard that it is raining from the weather forecast or from a friend.

(9)   pi-ka   o-n-tay
       rain-NOM   come-PRES-HEARSAY.EV
       a. 'It is raining (, I heard on the weather forecast.)'
       b. 'It is raining (, I heard from Kim.)'

The Korean inferential evidential *-napo-* signals that the information referred to in the utterance comes from the speaker's inference process, and therefore implies that the speaker is not certain

about *p* (e.g. Kwon, 2012). (10) can be uttered in a situation where the speaker only infers that it is raining, without any direct perceptual evidence (e.g. when the speaker sits in a windowless room and sees people walk in with wet umbrellas.)

(10) pi-ka   o-napo-a
      rain-NOM   come-INF.EV-DECL
      'It is raining (, I infer.)'

It is important to note that all three of these evidential markers can felicitously occur in sentences with both PPTs and MDs, as will become clear in examples (13) and (14).

4. Aims of the present work

The main aim of this paper is to experimentally investigate how evidentiality and subjectivity interact. We particularly focus on the role of *direct experience* in guiding the perceived generalizability of opinions. We test whether the presence of direct evidence for the subjective opinion – as signaled by (i) direct evidential marking and/or by (ii) the first-person experience requirement of PPTs – influences the perceived generalizability of opinions.

Given claims that firsthand experience is more reliable than indirect evidence (e.g. de Haan, 1998; Willett, 1988; Faller, 2002; Davis, Potts, & Speas, 2007), one might hypothesize that direct evidential marking boosts the perceived generalizability of opinions. Psychological work showing that direct first-person experience has a privileged cognitive status also supports this. For example, first-person experience has a stronger effect on attitudes and behavior than less direct information (e.g. Regan & Fazio, 1977; Fazio & Zanna, 1981; DeLamater et al., 2014).

Thus, under the **Direct Experience Advantage Hypothesis,** direct evidential marking (signaling first-person experience) increases the perceived generalizability of opinions: Encountering an opinion marked with a direct evidential leads comprehenders to perceive that opinion as one that many people would agree with, given that the speaker has direct, reliable evidence for her view. Therefore, subjective opinions marked with direct evidentials are predicted to elicit higher levels of perceived generalizability than opinions marked with hearsay and inferential evidentials. This is illustrated in (11). We do not assume any specific ordering of hearsay and inferential evidentials, given claims that the relative ordering between the two is context-dependent (e.g. Faller, 2002; Davis, Potts, & Speas, 2007; Papafragou et al., 2007).

(11)   Predictions of the Direct Experience Advantage Hypothesis (for PPTs and MDs)
       Direct evidential > {Hearsay evidential, Inferential evidential}

One may further hypothesize the 'direct evidence' generalizability boost to be clearer with PPTs than MDs, given claims that PPTs' semantics inherently rely on first-person experience, while opinions expressed with MDs can be formed based on second-hand reports and inferential evidence (see Section 2.1). Intuitively, the idea is that an opinion such as "the cake is tasty" would be perceived as more generalizable (held by a larger number of people) when it comes from someone who has tasted the cake, vs. someone who only has indirect evidence (inferential or hearsay evidence) about the tastiness of the cake. With MDs, such as "Kim is intelligent", the advantage of direct evidence may be weaker: A speaker may conclude from

her direct interactions with Kim that Kim is intelligent, but someone else with indirect evidence of Kim's intelligence (e.g. Kim's academic records) can have equally good (or even better) grounds for holding this opinion. Thus, effects of evidentiality might be less clear with MDs.

Alternatively, for PPTs in particular, it could turn out that – due to their direct experience requirement – this specific class of predicates is *not* susceptible to effects of different evidential markers. Under this view, the requirement that use of a PPT requires the judge to have direct first-person experience blocks potential modulating effects of evidential markers. We call this the ***Direct Experience Penalty Hypothesis for PPTs.*** The predictions are in (12).

(12)   Predictions of the Direct Experience Penalty Hypothesis (for PPTs)
       Direct evidential = Hearsay evidential = Inferential evidential

According to this hypothesis, whether opinions expressed with PPTs – based on *first-person experience* – are marked with a direct, indirect or hearsay evidential is irrelevant for perceived generalizability. At first glance, this may appear counterintuitive. But let's take a closer look at the judge=experiencer requirement of PPTs. It means that a non-experiencer individual is typically not entitled to be the judge of a PPT (unlike with MDs). That is, one is usually not entitled to utter a sentence like "The cake is tasty" merely on the basis of having heard someone else's description of the cake (hearsay evidence), or on the basis of having seen another person eat multiple servings of cake (inferential evidence). In these cases, it would be more appropriate to say "I heard that the cake is tasty" or "Apparently, the cake is tasty." Thus, regardless of how "reliable" a speaker's account of a cake's taste is (whether it's marked with a direct, hearsay or inferential evidential), it cannot be generalized straightforwardly to others because the others *need to experience it themselves* in order to be able to form their own opinions. This line of reasoning yields the prediction that the perceived generalizability of opinions expressed with PPTs is not modulated by evidential strength, as in (12). (This second hypothesis only applies to PPTs, not MDs, because MDs do not have the judge=experiencer requirement.)

5. Experiments

To assess the validity of the Direct Experience Advantage Hypothesis and the Direct Experience Penalty Hypothesis, we conducted two web-based studies using a method similar to Kaiser & Lee (2019, 2021): We used an alien planet context where participants read opinions expressed with PPTs (Experiment 1) or MDs (Experiment 2) and provided a numerical response indicating how many aliens they think would also hold this opinion. Because Experiments 1 and 2 had a parallel design and procedure, and differed only in terms of the adjectives, we describe the methods jointly for both experiments in Section 5.1.

5.1 Methods

5.1.1 Participants

Participants, most of whom were residing in Korea, participated remotely over the internet (73 participants in Experiment 1, PPTs, and 89 in Experiment 2, MDs). All were self-reported native speakers of Korean with normal or corrected-to-normal hearing and vision. At the end,

participants had the opportunity to enter a raffle to receive a gift card. Given the potential for distraction with remote web-based studies – and in light of recent work showing that data quality on web-based platforms such as Amazon Mechanical Turk has decreased strikingly (e.g. Chmielewski & Kucker, 2020) – we decided beforehand to use both practice items and catch trials (attention checks during the experiment) to identify who understood the task correctly and paid attention. We decided (before data analysis) to only include people who made no errors on five unambiguous practice items and who made no more than one error on six unambiguously true or false catch trials during the experiment. (e.g. "The first syllable of the word *mok-ley-seong* is *mok*" (true), "*peot* is made up of two syllables." (false)).

The exclusion criteria were pre-specified before data analyses on target trials were conducted, and lead to 24 participants being excluded in Experiment 1 and 40 in Experiment 2. While these exclusion rates may seem high, they are in line with what has been observed for web-based studied (see e.g. Chmielewski & Kucker, 2020, who observed 38-62% of MTurk participants failing at least one data quality validity indicator in summer/fall 2018 and spring 2019). Our use of multiple measures (practice items and catch trials) to assess whether participants were attending to the task fits with the advice of Chmielewski and Kucker (2020). The final analyses for both experiments included 49 native Korean speakers per experiment.

5.1.2 Materials and design

Both Experiments 1 and 2 consisted of 24 targets, 36 fillers, and 6 catch trials. Half of the critical adjectives in each study were positive (e.g. *caymi-iss-nun* 'fun', *ywuyonghan* 'useful') and half were negative (e.g. *cilwuhan* 'boring', *ssultey-eps-nun* 'useless'). The proportion of negative and positive adjectives in our experimental conditions was counterbalanced. Target items were in colloquial direct speech form and presented as being said by an alien. Following Kaiser and Lee (2019, 2021), we chose an alien planet context with nonce nouns in order to avoid people's own opinions about actual things influencing their ratings of opinion generalizability. The critical subjective predicate was in predicative position, and described a nonce noun. The nonce nouns were pronounceable nonwords of Korean (e.g. *pulon*, *letum*, *kumwupa*). In the sentence said by the alien, we manipulated the type of evidentiality marking. Examples are in (13) and (14). We tested four conditions: (a) no evidentiality marking (baseline), (b) direct evidential, (c) inferential evidential, (d) hearsay evidential. After each sentence, participants answered a question (e.g. (15)) about how many aliens would have the opinion expressed in the critical sentence (e.g. the pulon is tasty, the kumwupa is useless), out of 100 randomly-selected aliens on this planet. Crucially, the question itself was not marked with an evidential morpheme.

(13) Example stimuli for Experiment 1 (PPTs)
      Suppose you are visiting an alien planet, and hear an alien say:
      a. Baseline
         pulon   ku-ke   massis-e
         pulon   that-thing   tasty-DECL
         'The pulon is tasty.'
      b. Direct evidential (*-te-*)
         pulon   ku-ke   massis-te-la
         pulon   that-thing   tasty-DIR.EV-DECL

      'The pulon is tasty (, I have direct perceptual experience.)'
  c. <u>Inferential evidential (*-napo-*)</u>
     pulon   ku-ke   massis-napo-a
     pulon   that-thing   tasty-INF.EV-DECL
     'The pulon is tasty (, it seems.)'
  d. <u>Hearsay evidential (*-tay*)</u>
     pulon   ku-ke   massis-tay
     pulon   that-thing   tasty-HEARSAY.EV
     'The pulon is tasty (, I hear.).'

(14) Example stimuli for Experiment 2 (MDs)
     Suppose you are visiting an alien planet, and hear an alien say:
     kumwupa   ku-ke   ssulteyeps-{(*a*) e/(*b*) te-la/(*c*) napo-a/(*d*) -tay}
     kumwupa   that-thing   useless-{DECL/DIR.EV-DECL/INF.EV-DECL/HEARSAY}
     'The kumwupa is useless ({Ø/I have direct perceptual experience/it seems/I hear.})'

(15) Example questions
     If we select 100 random aliens that live on this planet,
  a. pulon-i   massis-ta-nun   uykyen-ul   kaciko iss-ulkka-yo?
     pulon-NOM   tasty-DECL-ADJ   opinion-ACC have   exist-Q-HON
     <u>'how many of them would have the opinion that the pulon is tasty?</u>'   [PPT]
  b. kumwupa-ka   ssulteyeps-ta-nun   uykyen-ul   kaciko iss-ulkka-yo?
     kumwupa-NOM   useless-DECL-ADJ   opinion-ACC have   exist-Q-HON
     <u>'how many of them would have the opinion that the kumwupa is useless</u>?'[MD]

Experiment 1 tested 24 PPTs and Experiment 2 tested 24 MD predicates; each study had 24 targets (plus 36 fillers and 6 catch trials), presented using a standard Latin-Square design. Thus, each participant only saw each PPT and each MD once, and did not see more than one version of each item. The only difference between the studies are the predicates in the targets; the design, procedure, and other properties of the targets and fillers are otherwise the same in both.

In the target items, *ku-ke* 'that thing' was inserted to avoid the subject noun (e.g. *pulon*, *kumwupa*) being interpreted as the experiencer. In Korean, sentences like "*emma* ('mom') *massis-napo-a*", for example, can mean 'Mom finds it tasty (, I infer)', an interpretation available if it is assumed that the nominative case marker and the object noun are omitted (which is common in a language like Korean where case-drop and object-drop are frequent). Thus, "pulon massis-napo-a", for instance, can receive a reading like 'pulon (an alien's name) finds it tasty', where the nonce noun *pulon* is interpreted as a sentient experiencer, not the thing being experienced/evaluated. This is not a reading that we intend. Inserting *ku-ke* 'that thing' disambiguates the nonce noun *pulon* to only refer to a non-human thing being experienced/evaluated (and bans the alien name reading). Thus, all predicates describe non-human objects.

Both experiments contained 36 fillers. Fillers were aimed to elicit a wide range of numerical responses to encourage participants to use the full 0-100 scale. Some asked about how many aliens would know something (objective fact); others asked about how many aliens would agree with something (more subjective). Some contained nonce words. Some were marked

with the direct evidential marker, the inferential evidential marker, or the hearsay evidential marker. The fillers in Experiments 1 and 2 were the same.

5.1.3 Procedure

Participants completed the experiment on the Qualtrics web interface. In both experiments, participants were provided with an alien planet context, similar to the context in Kaiser & Lee (2019, 2021). The context says: "Imagine that you are visiting an alien planet, and hear an alien say the following:" After the provided context, participants saw the alien's utterance containing a PPT (Experiment 1) or an MD (Experiment 2), marked with parentheses. Then, the participants were asked "If we select 100 random aliens that live on this planet, how many of them would have the opinion that the [nonce noun] is [PPT/MD]?" The task was to type in a numerical response (between 0 and 100) in a text box below the question. The context sentence, the PPT/MD sentence, and the question were all presented on the same screen.

5.2 Predictions

According to the *Direct Experience Advantage Hypothesis,* direct evidential marking boosts the perceived generalizability of opinions with both PPTs and MDs. Thus, this hypothesis predicts that the direct evidential condition will elicit higher levels of perceived generalizability (i.e. more aliens who have the opinion expressed in the critical sentence) than the hearsay or the inferential evidential conditions, which are not predicted to differ from each other: *direct > {hearsay, inferential}*. This asymmetry be may clearer with PPTs than with MDs, given that opinions expressed with PPTs require first-person experience, while opinions expressed with MDs can also be formed based on second-hand reports and inferential evidence.

In contrast, the *Direct Experience Penalty Hypothesis*, which only applies to PPTs, predicts that there will be no effect of evidential type on the perceived generalizability of opinions expressed with PPTs. In other words, with PPTs, all four evidentiality types should yield comparable numbers of aliens perceived to have the opinion (*direct=hearsay=inferential).* When considering this prediction, it is important to keep in mind that the critical question (e.g. (15)) had no evidential marking (e.g. *How many aliens would have the opinion that the pulon is tasty?*). Thus, in order for other aliens to have this opinion, they must have direct experience of tasting the pulon. In the context of our experiment, this information is left underspecified. Thus, participants are free to imagine a context where no other aliens have not had the relevant kind of direct experience, in which case generalization is not possible, regardless of the type of evidence. (Recall that PPTs are unlike MDs in having this specific first-person experience requirement; opinions expressed with MDs can be formed based on a much broader range of experiences and types of evidence.) In sum, this hypothesis therefore predicts that with PPTs, all four conditions will pattern similarly. Evidential marking does not affect what participants know about the evidential grounds of *other aliens*.

As shown in (13) and (14), we also included a baseline condition, with no evidential marking. However, this is largely an exploratory condition, because the two hypotheses that we test, the *Direct Experience Advantage Hypothesis* and the *Direct Experience Penalty Hypothesis,* make no specific predictions about the baseline. It could be that an opinion with no evidential

marking is perceived as the most generalizable, because one might think that the speaker is so very certain of her opinion that she does not even need to signal her source of evidence to back it up. Alternatively, the direct evidential condition could be perceived as more generalizable than the baseline, because the presence of the direct evidential signals more reliable evidence.

5.3 Results

For the response data, we analyzed both the raw data (numerical responses from 0 to 100) and z-scores. We report the results for raw responses; the z-score analyses showed similar results. Figure 1 shows the average response for each condition, for each experiment. In other words, how many aliens did participants report as having the same opinion? For data analysis, we used linear mixed-effects models. We used the lme4 package (Bates et al., 2015) in the R software environment (R Development Core Team, 2019). In order to control for potential effects of predicate valence, we included valence as a fixed effect. We deviation-coded evidential type and included it as a fixed effect. Specifically, we used deviation coding adjusted for four levels (k-1/k for the reference level and -1/k for others). The direct evidential condition was first set as the reference level and compared to the other three conditions (direct vs. hearsay, direct vs. inferential, direct vs. baseline). Then, the reference level was changed to the hearsay evidential condition to compare hearsay vs. inferential and hearsay vs. baseline. Lastly, the reference level was set to the inferential evidential condition to conduct the inferential vs. baseline comparison. For all analyses, we used the maximal model structure supported by the data.

5.3.1 Results for Experiment 1: Predicates of personal taste

As shown in the top half of Figure 1 (PPTs), the four conditions pattern alike: When asked how many other aliens have the opinion expressed in the critical sentence (e.g. *the pulon is tasty*), participants' responses do not show clear sensitivity to evidential marking – all bars are about the same length. This is confirmed by statistical analysis: there is no effect of evidentiality type (p's > 0.2, t's < 1.3). (Despite appearances, the direct evidential condition does not differ significantly from others.) Participants perceive opinions expressed with PPTs to be equally generalizable regardless of whether they are marked with the direct evidential, the hearsay evidential, inferential evidential or no evidential. This is in line with the Direct Experience Penalty Hypothesis, but is not predicted by the Direct Experience Advantage Hypothesis.

5.3.2 Results for Experiment 2: Multidimensional adjectives

The bottom half of Figure 1 shows the results for the MDs. Visual inspection shows that the inferential evidential condition bar is lower (perceived as less generalizable) than the others. Statistical analyses confirm that responses in the inferential evidential condition are lower than those in the other three (hearsay evidential, direct evidential, and baseline (p's < 0.01, t's > 1.7; the direct evidential vs. inferential evidential comparison is significant with z-scores ($p=0.02$, $t=2.34$), but marginal with raw responses ($p=0.07$, $t=1.77$)). The finding that the inferential evidential condition elicits lower responses than other conditions (including direct evidentials) partly supports the Direct Experience Advantage hypothesis. Unexpectedly, there no clear difference between the hearsay and direct evidential conditions. We return to this in the general discussion.
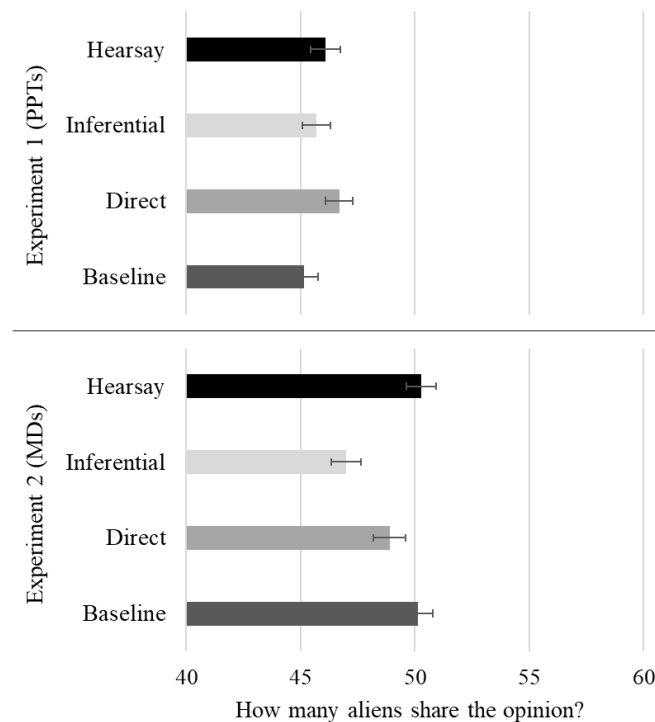
Figure 1. Opinion generalizability: How many aliens out of 100 share the opinion expressed in the critical sentence? (Error bars show +/- 1 SE)

5.3.3 Comparing Experiments 1 and 2

Opinions expressed with PPTs (Experiment 1) are perceived as numerically less generalizable than opinions expressed with MDs (Experiment 2). However, this difference is not statistically significant ($p > 0.1$, $t=0.9$). Although the lack of a significant difference precludes any firm conclusions, it is worth noting that this numerical pattern is in line with the Direct Experience Penalty Hypothesis. However, because PPTs and MDs were tested in different experiments, it is difficult to compare across the two. We leave this issue for future investigation.

6. General discussion

This paper investigates the relationship between subjectivity and evidentiality (see also Nuyts, 2001; Korotkova, 2016; Anand & Korotkova, 2018; Muñoz, 2020, i.a.). In particular, we explored the interpretation of PPTs (e.g. *tasty*) and non-PPT multidimensional adjectives (MDs, e.g. *useful*). It is widely agreed that the judge of PPTs needs the relevant kind of direct, first-person experience, but MDs are not subject to this requirement (e.g. Bylinina, 2014; McNally & Stojanovic, 2017). However, it has also been claimed that PPTs can express claims about people in general, thereby going beyond the first person (Moltmann, 2010; Pearson, 2013). The present work builds on our earlier studies on the generalizability of opinions in English (Kaiser & Lee, 2019; 2021) by extending the domain of inquiry to how *evidentiality* guides the perceived generalizability of opinions expressed with PPTs and MDs in Korean.

To assess how evidential cues – in particular direct experience, either signaled by the direct evidential or imposed as a semantic requirement on PPTs – guide the perceived generalizability

of opinions, we tested the *Direct Experience Advantage Hypothesis*, according to which the greater reliability of the direct evidential marking boosts the perceived generalizability of opinions with both PPTs and MDs, and the PPT-specific *Direct Experience Penalty Hypothesis*, according to which PPT opinions are not influenced by evidentiality.

Indeed, with PPTs we find no evidence of any evidentiality effects, which supports the *Direct Experience Penalty Hypothesis* for this specific class of adjectives. We view the lack of evidentiality effects with PPTs as stemming from the fact that non-experiencers cannot be legitimate judges, i.e., opinions expressed with PPTs are difficult to generalize *regardless of how reliable the evidential marking is.* (Admittedly, a null result is always difficult to interpret, but the fact that we *did* obtain significant results with MDs, using the same kind of design, task, and sample size, provides some indication that the lack of significant evidentiality effects with PPTs is unlikely to be a Type II error.)

In contrast to PPTs, with MDs we find that evidentiality does modulate the generalizability of opinions, partly in line with the *Direct Experience Advantage hypothesis*. Opinions marked with an inferential evidential are perceived as less generalizable than those marked with the direct evidential, hearsay evidential, or no evidential. That is, opinions that are rooted in inferential evidence are less generalizable than opinions based on direct or hearsay evidence. This finding fits with what is known about inference – that it signals a degraded level of certainty (e.g. Kwon, 2012). If someone infers something to be useful, this opinion is perceived as less "strong" than when someone who actually used it finds it useful. What is interesting in our data is that opinions marked with the hearsay evidential were perceived as *more* generalizable than opinions marked with the inferential evidential. This may be due to the hearsay evidential *-tay* being underspecified about its source (Section 3.3): If participants interpret *-tay* sentences to mean 'It is said that…' (indicating some level of social consensus), it is reasonable that they understood the opinion to be more generalizable. This is an important direction for future work.

The outcomes of our experiments point to a link between the perceived generalizability of opinions and the judge=experiencer requirement on a certain class of subjective adjectives (PPTs). Our results suggest that in order to answer the question of whether an opinion can be generalized to other potential judges, one needs to consider the evidential grounds on which the opinion is based. These results can provide a foundation for future investigations on a variety of topics, including steps to quantify differences in the level of communicability of individual adjectives, the differential interpretational role of hearsay evidential marking regarding social consensus for PPTs and MDs, etc. We also suggest that further research is needed to characterize the notion of evidential strength in the domain of statements that convey subjective opinion. The present work offers one potential way of approaching this, in terms of perceived opinion generalizability, but further investigation is needed. Finally, the finding that the evidentiality effect is modulated by the type of subjectivity (e.g. whether the opinion contains a PPT or an MD) suggests that there is much to be gained by further exploring the link between evidentiality and subjectivity.

References

Aikhenvald, A. (2004). *Evidentiality*. Oxford: Oxford University Press.

Anand, P. (2009). Kinds of taste. Manuscript. UCSC.

Anand, P., & Korotkova, N. (2018). Acquaintance content and obviation. In *Proceedings of Sinn und Bedeutung 22*. pp. 55–72.

Bates, D., M. Maechler, B. Bolker, and S. Walker. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1).

Bhatt, R. and R. Izvorski. (1998). Genericity, implicit arguments and control. *Paper presented at the 7th Student Conference in Linguistics*.

Bierwisch, M. (1989). The Semantics of Gradation. In M. Bierwisch and E. Lang (Eds), *Dimensional Adjectives*, pp. 71–261. Berlin: Springer-Verlag.

Bylinina, L. (2014). *The Grammar of Standards: Judge-Dependence, Purpose-Relativity, and Comparison Classes in Degree Constructions*. Utrecht: LOT.

Bylinina, L. (2017). Judge-dependence in degree constructions. *Journal of Semantics*, *34*(2), 291–331.

Bylinina, L., E. McCready, and Y. Sudo (2014), The Landscape of Perspective Shifting. *Talk given at Pronouns in Embedded Contexts at the Syntax-Semantics Interface*, University of Tübingen.

Chafe, W. L. and J. Nichols. (1986). *Evidentiality: The linguistic coding of epistemology*. Norwood, NJ: Ablex.

Chmielewski, M. and S. C. Kucker. (2020). An MTurk crisis? Shifts in data quality and the impact on study results. *Social Psychological and Personality Science*, *11*(4), 464–473.

Chung, K.-S. (2005). Space in tense: The interaction of tense, aspect, evidentiality, and speech act in Korean. Ph. D. Dissertation. Simon Fraser University.

Chung, K.-S. (2006). Korean evidentials and assertion. In *Proceedings of the 25th West Coast Conference on Formal Linguistics*, pp. 105–113.

Chung, K.-S. (2007). Spatial deictic tense and evidentials in Korean. *Natural Language Semantics*, *15*, 187–219.

Chung, J. Y. (2009). A dynamic semantic account of a family of Korean hearsay evidentials. Paper presented at the International Symposium on Korean Linguistics.

Collins, J. (2013). The syntax of personal taste. *Philosophical Perspectives*, *27*(1), 51–103.

Coppock, E. (2018). Outlook-based semantics. *Linguistics and Philosophy*, *41*(2), 125–164.

Davis, C., C. Potts, and M. Speas. (2007). The pragmatic values of evidential sentences. In *Proceedings of Semantics and Linguistic Theory (SALT) 17*, pp. 71–88.

DeLamater, J. D., D. J. Myers, and J. L. Collett. (2014). *Social Psychology*. Westview Press.

Faller, M. (2002). Semantics and pragmatics of evidentials in Cuzco Quechua. Ph. D. dissertation. Stanford.

Fazio, R. H., and M. P. Zanna. (1978). On the predictive validity of attitudes: The roles of direct experience and confidence. *Journal of Personality*, *46*(2), 228–243.

Fazio, R. H., D. M. Sanbonmatsu, M. C. Powell, and F. R. Kardes. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology*, *50*(2), 229.

de Haan, F. (1998). The category of evidentiality. Manuscript. University of New Mexico.

Kaiser, E. and J. Herron Lee. (2017). Experience matters: A psycholinguistic investigation of predicates of personal taste. In *Proceedings of Semantics and Linguistic Theory (SALT) 27*, pp. 323–339.

Kaiser, E. and J. Herron Lee. (2018). Predicates of personal taste and multidimensional adjectives: An experimental investigation. In *Proceedings of WCCFL 35*, pp. 224–231.

Kaiser, E. and S. H. Lee. (2019). On the generalizability of subjective opinions: Predicates of personal taste and multidimensionality. *Talk presented at Generative Linguistics in the Old World (GLOW) 42*, University of Oslo, Norway.

Kaiser, E. and S. H. Lee. (2021). What influences the perceived generalizability of opinions? Insights from subjective adjectives in non-egocentric contexts. Manuscript, University of Southern California.

Kaiser, E. and D. Rudin. (2020). When faultless disagreement is not so faultless: What widely-held opinions can tell us about subjective adjectives. In *Proceedings of Linguistic Society of America 5*(1), 698–707.

Kamp, H. (1975). Two theories about adjectives. In E. L. Keenan (Ed.), *Formal semantics for natural language*, pp. 123–155. Cambridge: Cambridge University Press.

Kennedy, C. (2013). Two sources of subjectivity: Qualitative assessment and dimensional uncertainty. *Inquiry*, *56*(2–3), 258–277.

Kim, Y.-H. (1981). Hoysangmwunuy Inching Ceyyakkwa Chaykimseng. (Person restriction and responsibility in recollecting speech). *Kukehak (Korean Linguistics)*, *10*, 37–80.

Kennedy, C. and M. Willer (2016). Subjective attitudes and counterstance contingency. In *Proceedings of the Semantics and Linguistic Theory Conference (SALT) 26*, pp. 913–933.

Klecha, P. (2014). Diagnosing modality in predictive expressions. *Journal of Semantics*, *31*(3), 443–455.

Klein, E. (1980). A semantics for positive and comparative adjectives. *Linguistics and Philosophy*, *4*(1), 1–45.

Korotkova, N. (2016). Disagreement with evidentials: A call for subjectivity. In J. Hunter, M. Simons, and M. Stone (Eds.), *JerSem: The 20th Workshop on the Semantics and Pragmatics of Dialogue*, pp. 65–75.

Kölbel, M. (2004). Faultless Disagreement. In *Proceedings of the Aristotelian Society 104*, pp. 53–73.

Kwon, I, (2009). The Korean evidential marker -te- revisited: its semantic constraints and distancing effects in mental spaces theory. *Paper Presented at Frames and Constructions*, University of California, Berkeley.

Kwon, I. (2012). Please confirm what I inferred: On the Korean inferential-evidential marker-napo. *Journal of Pragmatics*, *44*(8), 958–969.

Lasersohn, P. (2005). Context dependence, disagreement, and predicates of personal taste. *Linguistics and Philosophy*, *28*(6), 643–686.

Lasersohn, P. (2009). Relative truth, speaker commitment, and control of implicit arguments. *Synthese*, *166*(2), 359–374.

Lee, J. (2010). Evidentiality and its Interaction with Tense: Evidence from Korean. Ph.D. Dissertation. Ohio State University.

Lee, J. (2013). Temporal constraints on the meaning of evidentiality. *Natural Language Semantics*, *21*(1), 1-41.

Lim, D. (2010). Evidentials and Interrogatives: A Case Study from Korean. Ph.D. Dissertation. University of Southern California.

MacFarlane, J. (2006). Epistemic modals are assessment-sensitive. Manuscript, UC Berkeley.

McNally, L. and Stojanovic, I. (2017). Aesthetic adjectives. In J. Young (Ed), *The Semantics of Aesthetic Judgment*, pp. 17–37. Oxford: Oxford University Press.

Moltmann, F. (2006). Generic one, arbitrary PRO, and the first person. *Natural Language Semantics*, *14*(3), 257–281.

Moltmann, F. (2010). Relative truth and the first person. *Philosophical studies*, *150*(2), 187-220.

Muñoz, P. (2020). Acquaintance inference as evidential effects. Manuscript. University of Chicago.

Ninan, D. (2014). Taste predicates and the acquaintance inference. In *Proceedings of Semantics and Linguistic Theory (SALT) 24*, pp. 290–309.

Nuyts, J. (2001). Subjectivity as an evidential dimension in epistemic modal expressions. *Journal of Pragmatics*, *33*(3), 383–400.

Papafragou, A., P. Li, Y. Choi, and C. H. Han. (2007). Evidentiality in language and cognition. *Cognition*, *103*(2), 253–299.

Pearson, H. (2013). A judge-free semantics for predicates of personal taste. *Journal of Semantics*, *30*(1), 103–154.

R Core Team. (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/

Regan, D. T. and R. Fazio. (1977). On the consistency between attitudes and behavior: Look to the method of attitude formation. *Journal of Experimental Social Psychology*, *13*(1), 28–45.

Rudolph, R. (2019). Talking about Appearances: Experience, Evaluation, and Evidence in Discourse. Ph. D. Dissertation. University of California, Berkeley.

Sassoon, G. W. (2013). A Typology of Multidimensional Adjectives. *Journal of Semantics*, *30*(3), 335–380.

Seo, J.-S. (1977). 'te'nun hoysanguy kinungul cininunka. (Does 'te' have a retrospective function?). *Ene (Language)*, 2-1.

Snyder, E. (2013). Binding, genericity, and predicates of personal taste. *Inquiry*, *56*(2-3), 278–306.

Song, J.-M. (1998). Anmaycumssikkuth '-te-'uy uymi kinungey tayhaye: yuhyeonglon-cek kwancemeyse. (On the semantic functions of the imperfective marker '-te-': from a typological perspective). *Kukehak (Korean Linguistics)*, 135–169.

Song, J.-M. (2002). Typological analysis of the Korean evidential marker '-te-.' *Enehak (The Linguistic Society of Korea)*, *32*, 147–164.

Stephenson, T. (2007a). Judge dependence, epistemic modals, and predicates of personal taste. *Linguistics and Philosophy*, *30*(4), 487–525.

Stephenson, T. (2007b). *Towards a theory of subjective meaning*. Ph. D. Dissertation. MIT.

Stojanovic, I. (2007). Talking about taste: Disagreement, implicit arguments, and relative truth. *Linguistics and Philosophy*, *30*, 691–706.

Vardomskaya, T. (2014). Subjective adjectives as a cooperation problem. In *Proceedings of ConSOLE XXII*, pp. 191–204.

Willer, M. and C. Kennedy. (2020). Assertion, expression, experience. *Inquiry*, 1–35.

Willett, T. (1988). A cross-linguistic survey of the grammaticization of evidentiality. *Studies in Language*, *12*(1), 51–97.

Yang, I.-S. (1972). *Korean Syntax: Case Markers, Delimiters, Complementation, and Relativization*. Baykhap Publishing Co.