The perils of combining translation tasks and judgment tasks *

Katie Sardinha
Independent Researcher

Abstract: The *combination translation/judgment task* involves a translation task immediately followed by a judgment task, where the sentence to be judged is a minimally altered version of the sentence just volunteered in the translation task. While this composite task can be an efficient way of generating semantic data, it also has some significant vulnerabilities. In this paper, I examine transcribed dialogues from elicitation with four Kwak'wala language consultants and discuss how the structure of the combination translation/judgment task can give rise to an implication that the consultant's volunteered sentences are deviant. This implication, in turn, can undermine consultants' confidence in their own native speaker intuitions, which can change the way they respond to judgment tasks. To prevent this issue, careful attention must be paid to how the combination translation/judgment task is explained to and understood by consultants.

Keywords: translation task, judgment task, metalinguistic awareness, task descriptions, Kwak'wala

1 Introduction

Two of the most fundamental kinds of methods in semantic fieldwork are translation tasks and judgment tasks (Matthewson 2004; Bohnemeyer 2015). The purpose of this paper is to describe one way of efficiently combining these methods, as well as to discuss a significant but preventable problem that can result from this combination.

The variety of translation task that will concern us here is illustrated in (1). The elicitor ('KS') first defines a discourse context and then produces a sentence in the metalanguage which fits the given context. The language consultant ('Speaker') is then prompted to translate this sentence into the object language. In (1), the metalanguage is English and the object language is Kwak'wala, a Wakashan language spoken on the central coast of British Columbia.

(1) Translation task
   a. KS: Let's say Simon, he's outside and he's playing around with a ball, kicking it around.
   b. Speaker: Mhm.
   c. KS: [...][1] And then, let's say you're watching him and he, you know, he kicks at the ball, but he misses it. How would you say, 'He kicked at the ball, but he missed it'...?

---

[1]Square brackets containing an ellipsis ('[...]') indicate places where short, irrelevant stretches of speech have been omitted from the transcription presented here. Square brackets following a Kwak'wala sentence and containing a letter (e.g. '[e]' in 1d) point to the line within the same numbered example where the sentence is glossed.

---

d. Speaker: x̌iqʷi Simon la?i q̓əy̓ax?id x̌a luqʷsəm q̓əy̓akasu?. [e]

e.
x̌iqʷi			Simon	la?i			q̓əy̓ax?id		x̌a		luqʷsəm	q̓əy̓akasu?
x̌iqʷ	=i		Simon	lə	=a	=i	q̓əy̓ak	-x?id	=x̌	=a	luqʷsəm	q̓əy̓ak
miss	=D3		Simon	AUX	=EMB	=D3	kick	BEC	=ACC	=D4	round	kick

-a	-su?
-A	-ACC.PASS

lit. 'Simon missed when he kicked at the ball.'[2]   (20160713-S2 VF)

Translation tasks are one way of generating positive evidence for a semantic hypothesis. The example in (1), for instance, was elicited to see whether the direct object of the verb *q̓əy̓ak-* 'kick' can be accusative when the predicate has a connative interpretation (it can).

In a judgment task, the elicitor defines a discourse context and then asks the consultant to judge whether a grammatical sentence in the object language is semantically acceptable in that context. From the analyst's perspective, an 'acceptable' sentence is one that expresses a proposition that is both true and felicitous in the context. From the language consultant's perspective, what counts as an 'acceptable' sentence has a phenomenological basis: an acceptable sentence is one that sounds fitting, sensical, and natural in the given context. An example of a judgment task where the judged sentence is accepted is shown in lines (2h)-(2j) of (2). Since the discourse context for this judgment task involves a translation task (2e)-(2g), this portion of the dialogue is included as well.

(2) Judgment task (accepted)
    a. KS: So let's say Mabel's dad, Bill, is gonna have open heart surgery. And she's scared for him...

    b. Speaker: Mhm.

    c. KS: ...that it could go wrong.

    d. Speaker: Mhm.

    e. KS: How would I express, um, 'Mabel's scared for her dad'...?

    f. Speaker: I think *nuła* is 'concerned'. *kəłəla* is 'afraid' or 'scared'. *nuła* is 'concerned'. *nułux̌ Mabelx̌ qə?es ?ump*. [g]

    g.
nułux̌		Mabelx̌		qə?es		?ump
nuł	=ux̌	Mabel	=(ə)x̌	qa	=i(?)s	?ump
concerned	=D2	Mabel	=V2	PREP	=3.REFL.POSS[D3]	father

'Mabel is scared for her dad.'   (20160721-S2 VF)

    h. KS: [...][3]And does it make sense to say, um, *kəłəlux̌ Mabelx̌ qa?us ?ump* [i]...?

---

[2]Abbreviations used in the transcription and glossing of examples include the following: - 'suffix', = 'clitic', °'softening mutation', 2 'second person', 2.POSS'second person possessor', 3.REFL.POSS 'third person possessor coreferent with subject or topic', A 'default vowel', ACC 'accusative', ACC.PASS 'accusative passive', AUX 'auxiliary', BEC 'become operator', CONT 'continuative', D2 'third person medial deictic determiner', D3 'third person distal deictic determiner', D4 'existential deictic determiner', DIR.TEL 'telic directional', EMB 'embedding vowel', INST 'instrumental', JF 'judged form', lit. 'literally', PREP 'preposition', V2 'third person medial visible', VER 'verum focus', VF 'volunteered form'.

i. kəɬəlux̌   Mabelx̌   qaʔus   ʔump
    kəɬ   -əla =ux̌ Mabel =(ə)x̌ qa   =u(ʔ)s   ʔump
    scared -CONT =D2 Mabel =v2 PREP =3.REFL.POSS[D2] father
    'Mabel is scared for her dad.'   (20160721-S2 JF)

j. Speaker: Mhm. *ʔikʔəmx̌e*. ('That's good too.')

From the judgment task in (2), we learn something about the lexical semantics of the verb *kəɬəla* 'scared, afraid', namely that it can be used in the same context as another verb, *nuɬa*, to express concern for a person.

A sentence that is rejected in a judgment task is either false or infelicitous in the given context. Example (3) shows an example of a judgment task where the sentence in (3a)/(3g) is rejected. Note that rather than rejecting the sentence directly, the consultant does so indirectly by correcting the offered sentence to make it express the appropriate meaning in the context. It is often the case when doing judgment tasks that follow-up discussion is needed to discover why a particular sentence has been rejected.

(3) Judgment task (rejected)

    [Continuing to work within the context in (2a-c) where Mabel is scared for her Dad...]

    a. KS: And *ġiġeʔqəlux̌ Mabelx̌ sis ʔum*p...? [b]

    b. #ġiġeʔqəlux̌   Mabelx̌   sis   ʔump
        ġiġeʔq   -əla =ux̌ Mabel =(ə)x̌ =s   =i(')s   ʔump
        worry   -CONT =D2 Mabel =v2 =INST =3.REFL.POSS[D3] father
        lit. 'Mabel is afraid of her dad.'   (20160721-S2 JF)

    c. Speaker: *qaʔis ʔump.*

    d. KS: *qaʔis*?

    e. Speaker: Mhm.

    f. KS: Not *sis*?

    g. Speaker: *ġiġeʔqəlux̌ Mabelx̌ sis ʔump.* [b] That means that she's 'afraid of her dad', *sis ʔump.* But if you say *qaʔis ʔump*, 'for'.

Judgment tasks are useful for generating negative evidence for semantic hypotheses. The example in (3), for instance, provides evidence that a benefactive interpretation is not available for phrases headed by the instrumental case marker, *=s*.

One very efficient way of generating semantic data is to combine translation tasks and judgment tasks. What I will refer to as a *combination translation/judgment task* involves a translation task immediately followed by a judgment task, where the sentence to be judged is a minimally altered version of the sentence just volunteered in the translation task. Combination translation-judgment tasks involve the generalized procedure in (4):

---

[3]At this point in the dialogue, the Speaker volunteers a Kwak'wala sentence that elaborates on (2g), describing why exactly Mabel is scared for her Dad (*...qu laƛ qʷax̌ʔidsuʔƛ* 'for if he's gonna be operated on'). The judgment that immediately follows is understood as being within this same context.

(4) Combination translation/judgment task

 a. Define a discourse context.

 b. Carry out a translation task from the metalanguage into the object language, obtaining a volunteered translation in the object language.

 c. Take the volunteered translation from step (b) and alter it minimally in a way that is relevant to the hypothesis being investigated. The sentence created in this way is called the constructed sentence.

 d. Carry out a judgment task using the constructed sentence to see whether it is acceptable within the discourse context defined in (a).

A volunteered translation is altered 'minimally', as referenced in step (4c), if the volunteered translation and the constructed sentence are identical except for one change. For instance, the translation task in (2) and the follow-up judgment task in (3) do not qualify as a combined translation/judgment task because the difference between either (2g) or (2i) and (3b) is not minimal: both the verb and the way the internal argument of the verb is marked have been altered.

I have made extensive use of combination translation/judgment tasks to investigate the semantics of object case marking in Kwak'wala (Sardinha 2016, 2017, 2018). In this language, there are two morphological cases for marking objects: instrumental ($=s$) and accusative ($=\check{x}$)[4]. Empirically, investigating the semantic basis of object case has primarily involved looking at which contexts with transitive predicates allow instrumental objects only, which allow accusative objects only, and which allow objects in either case, in addition to looking at the range of interpretations that can be coerced by the presence of each case marker.[5]

An example of a combined translation/judgment task being used to investigate the semantics of object case in Kwak'wala is shown in (5). The consultant's volunteered sentence contains an accusative object, and the constructed sentence contains an instrumental object; otherwise, these sentences are identical. In this example, the consultant accepts the constructed sentence and comments that it sounds semantically equivalent to the translation she had just volunteered in the same context.

(5) Combination translation/judgment task (constructed sentence is accepted)

 a. KS: So let's say, um, Mabel – she runs into the room. She has this book. And she puts it down on the table quickly, and then she runs out. How would I say 'Mabel put down the book quickly'…?

 b. Speaker: *haṅakʷili Mabel ʔəx̌ʔaliɫa x̌a ḱaḱadəkʷsiladzu.* [c]

 c.

haṅakʷili		Mabel	ʔəx̌ʔaliɫa				x̌a		ḱaḱadəkʷsiladzu
haṅakʷila	=i	Mabel	ʔəx̌	-(g)aʔɫ	-°iɫ	-a	=x̌	=a	ḱaḱadəkʷsiladzu
do.quickly	=D3	Mabel	do	-DIR.TEL	-in.house	-A	=ACC	=D4	book

'Mabel put the book (ACC) down quickly (in the house).' (20160707-S2 VF)

---

[4]More specifically, the case markers $=s$ 'instrumental' and $=\check{x}$ 'accusative' are used for marking third person arguments. There is a separate set of forms for first person and second person objects (Sardinha 2017:6).
[5]See for instance sentence (6m), where instrumental case coerces a nonsense interpretation.

d. KS: I'm gonna change it a little bit, and you can tell me if it's...

e. Speaker: *ẃiga* ('Let's go.').

f. KS: *haṅakʷili Mabel ʔəx̌ʔaliła **S**A ḱaḱadək̓ʷsiladzu*. [g]

g.
haṅakʷili	Mabel ʔəx̌ʔaliła				**sa**		ḱaḱadək̓ʷsiladzu
haṅakʷila =i	Mabel ʔəx̌	-(g)aʔł	-°ił	-a	**=s**	=a	ḱaḱadək̓ʷsiladzu
do.quickly =D3	Mabel do	-DIR.TELL	-in.house	-A	**=INST**	=D4	book

'Mabel put the book (**INST**) down quickly (in the house).'   (20160707-S2 JF)

h. Speaker: Mhm. It's, it's just as legal as the first one.

i. KS: Kay. And do you notice any difference?

j. Speaker: It doesn't change what you're saying.

k. KS: Yeah. Does it change the way you're looking at it?

l. Speaker: Mhm, no.

m. KS: No.

n. Speaker: It's just the sound, that's all you've changed. But it's still, uh, understandable.[6]

Another example of a combination translation/judgment task used to investigate object case is shown in (6). In this example, the consultant's volunteered sentence contains an accusative object and the constructed sentence contains an instrumental object. This time the consultant rejects the constructed sentence, expressing this rejection by providing an obviously nonsensical translation for it. (This nonsense translation was accompanied by body language and a tone of voice that communicated disapproval of the constructed sentence, aspects of the situation that are not apparent from the transcript below).

(6) Combination translation/judgment task (constructed sentence is rejected)

a. KS: 'Katie is putting the hat away.' {*KS puts the hat she is holding into her cloth bag.*}

b. Speaker: Finally going to sleep.[7]

c. KS: Yeah.

d. Speaker: *ləṁux̌ gəʔx̌ux̌ Katiyəx̌ʷa ƛətəmł*. [i]

e. KS: *ləṁux̌ gəx̌ux̌...?*

f. Speaker: *gəʔx̌ux̌ Katiyəx̌ʷa ƛətəmł*.

g. KS: *...Katiyəx̌ʷa ƛətəmł*. Um,*ləṁux̌ gəʔx̌ux̌ Katiyəx̌ʷa ƛətəmł*. [i]

h. Speaker: *ʔəm* ('Yeah').

---

[6]In sentence (5g), the instrumental case adds meaning that is redundant with respect to entailments of the verb. Hence, the speaker's judgment in (5h) is a judgment about semantic equivalence between (5c) and (5g) and is not merely a grammaticality judgment. See Sardinha (2017:50-88) for relevant discussion.

[7]We had been talking about putting the hat places for quite some time already in the session, and Speaker 2 is making a joke here about how tired the hat must be after its journeying.

i. ləm̓ux̌   gəʔx̌ux̌   Katiyəx̌   **x̌ʷa**   λətəmɬ
    lə   =ʔm =ux̌ gəʔx̌ =ux̌ Katie   =(ə)x̌ =x̌a =ʷ =a λətəmɬ
    AUX =VER =D2 stow =D2 Katie   =V2   =ACC =D2 =D4 hat
    'Katie is putting the hat (**ACC**) away.'   (20160720-S2 VF)

j. KS: Can I say, *ləm̓ux̌ gəʔx̌ux̌ Katiyəx̌ sux̌ λətəmɬ* [m]...?

k. Speaker: 'You're using the hat to put the hat away'.[8]

l. KS: Kay. *k̓i's nəqa* ('Not correct').

m. #ləm̓ux̌   gəʔx̌ux̌ Katiyəx̌   sux̌   λətəmɬ
    lə   =ʔm =ux̌ gəʔx̌ =ux̌   Katie =(ə)x̌ **=s** =ux̌
    AUX =VER =D2 stow =D2   Katie =V2 =INST =D2
    lit. 'Katie is using the hat (**INST**) to put the hat away.'   (20160720-S2 JF)

Combining translation tasks and judgment tasks in the way illustrated in (5)-(6) is method-ologically efficient, as it allows the researcher to use one discourse context to obtain two types of data in rapid succession. This is very useful in any fieldwork context where elicitation time is limited. The combination translation/judgment task can also help reduce the number of contexts which need to be introduced in a session. This is helpful because introducing too many discourse contexts in one elicitation session can be mentally taxing for the language consultant.

Despite these advantages, combining translation tasks and judgment tasks in the way I've described is not always benign. In fact, a significant problem with this method became apparent to me in the course of my fieldwork on object case. In the remainder of this paper, I will examine how the combination translation/judgment task led Kwak'wala consultants to become metalingistically aware of the topic under investigation, and how the structure of the combination translation/judgment task resulted in several language consultants beginning to doubt the reliability of their own native speaker intuitions, leading them to change the way they responded to the judgment portion of the task. I will begin by examining the empirical pattern that led me to discover that there was a problem associated with the method (Section 3), and then will explain what I believe to be the nature and origin of this problem (Section 4). Following this, I will discuss how the problem may be prevented by carefully explaining the task to consultants (Section 5). In the final section (Section 6), I will discuss the relevance of these findings to semantic fieldwork more generally.

2   Detecting a problem

The fieldwork under discussion in this paper was carried out during the summer of 2016 in three BC communities: Fort Rupert, Tsulquate, and Vancouver. Six native Kwak'wala speakers, all in their mid-seventies to mid-eighties, worked as language consultants to investigate the semantics of object case marking in Kwak'wala. Of these six consultants, four completed combination translation/judgment tasks, while the remaining two consultants contributed positive data to the project but did not provide judgments. All of the elicitation sessions referenced below were carried out one-on-one. The four consultants who engaged in combination translation/judgment

---

[8]The interpretation of the hat as both instrument and theme in this example is notable, and is explained by a movement analysis in Sardinha (2017:170-175).

tasks, who are referred to as Speaker 1, 2, 3, and 4 below, had all previously worked with me for between five and seven years at the time of investigating object case. All were also involved, at least sporadically, in Kwak'wala language teaching, and all had experience working with linguists other than me. Significantly, all four of these consultants were experienced in carrying out both translation tasks and judgment tasks prior to engaging in combination translation/judgment tasks.

   I realized that the combination translation/judgment task was causing problems when these four consultants began exhibiting two different patterns of responses to the task. On the one hand, Speaker 1 and Speaker 2 appeared to be carrying out the combination translation/judgment task as intended. Their data exhibited the overall pattern in (7):

(7)   Combination translation/judgment task data: Speaker 1 and Speaker 2

   a.   Speaker 1 and Speaker 2's judgment data agreed with each others', but differed from that of Speaker 3 and Speaker 4.

   b.   Speaker 1 and Speaker 2's judgment data were consistent with their positive (corpus) data. For example, neither speaker was observed to spontaneously produce a sentence with the verb *kəɬ-* 'afraid' and an accusative object, and both speakers rejected sentences with *kəɬ-* and an accusative object in the judgment portion of the task.

   c.   Speaker 1 and Speaker 2 accepted some constructed sentences and rejected some constructed sentences.

   The semantic data from Speaker 3 and Speaker 4, on the other hand, indicated that these speakers were responding to the task in some unanticipated way. In contrast with Speaker 1 and Speaker 2, these speakers showed a strong tendency to accept every constructed sentence they were asked to judge. Their overall data exhibited the pattern in (8):

(8)   Combination translation/judgment task data: Speaker 3 and Speaker 4

   a.   Speaker 3 and Speaker 4's judgment data agreed with each others', but differed from that of Speaker 1 and Speaker 2.

   b.   Speaker 3 and Speaker 4's judgment data were not always consistent with their positive (corpus) data. For example, neither speaker was observed to spontaneously produce a sentence with the verb *kəɬ-* 'afraid' and an accusative object, though both speakers accepted sentences with *kəɬ-* and an accusative object in the judgment portion of the task.

   c.   With few exceptions, Speaker 3 and Speaker 4 accepted every constructed sentence.

   In the midst of elicitation, it was puzzling to me why these two language consultants were accepting nearly every constructed sentence when doing the combination translation/judgment task. As elicitation proceeded, however, clues about what was happening gradually emerged from these speakers' metalinguistic commentary.[9]

---

[9]These clues became more apparent during the process of transcribing these sessions, which took place a few months after the sessions themselves. Yet while it would take months studying the transcriptions of sessions for me to understand clearly what was happening with respect to the combination translation/judgment task,

The transcriptions of commentary in examples (9)-(12) are taken from elicitation with Speaker 3. In (9), which took place soon after we began doing combination translation/judgment tasks, Speaker 3 rejects a constructed sentence but mentions that other speakers might have a different judgment than her.

(9) Speaker 3 commentary: volunteered form with accusative object (=x̌), judged form with instrumental object (=s)

    a. Speaker: I wouldn't [put *sux̌* in the sentence]. But...

    b. KS: Okay.

    c. Speaker: ...some people might be different, hun.   (20160707-S3 VF)

This explicit focus on how other Kwak'wala speakers might judge the constructed sentence occurred several times during the same session. At one point in the session, some time after the dialogue in (9), KS clumsily tried to address Speaker 3's focus on other Kwak'wala speakers by expressing that she takes Speaker 3's native speaker intuitions seriously. In response, Speaker 3 reaffirms the possibility of other speakers' intuitions being different than hers (10).

(10) Speaker 3 commentary

    a. KS: When I ask you these questions, so if it's something you wouldn't say...

    b. Speaker: Uh-huh.

    c. KS: ...then, I'll, say that it's not good.

    d. Speaker: *ʔəm* ('Yes').

    e. KS: Cuz you're the expert.

    f. Speaker: Okay. But somebody might say different.   (20160707-S3 VF)

As this same session progressed and more combined translation/judgment tasks were carried out, Speaker 3 began not just accepting every constructed sentence, but even expressing a preference for each constructed sentence over the one she had just previously volunteered. This is reflected in the commentary in (11), in which Speaker 3 responds to a judgment task by reflecting negatively on the sentence she had just volunteered (which contained an accusative object). She then remarks that the constructed sentence she has been offered to judge (with an instrumental object) would be a more proper sentence for the purpose of teaching the language.

(11) Speaker 3 commentary: volunteered form with accusative object (=x̌), judged form with instrumental object(=s)

    a. Speaker: I think – sometimes mine are a little too... I think if you're teaching somebody, that second one you said [with =*sada*], sounds better.   (20160707-S3 VF)

---

it was obvious at the time of elicitation that the task was causing a major disruption to these two consultants' normal response patterns. As soon as I realized this, I stopped using the combination translation/judgment with these two consultants.

At another late point in the same session, Speaker 3 takes this self-criticism one step further and denigrates her own abilities as a speaker (12).

(12) Speaker 3 commentary: volunteered form with instrumental object (=*s*), judged form with accusative object (=*x̌*)

   a. Speaker: I think that's the, proper sentence [with =*x̌a*]. Hun. Not mine. I always forget a little something.   (20160707-S3 VF)

Comments like those in (9)-(12) suggest that something about the combination translation/judgment task is undermining Speaker 3's confidence in her own semantic intuitions, or at least her confidence in her intuitions being representative of the wider Kwak'wala speech community.

   Speaker 4's response to the combination translation/judgment task was somewhat different from Speaker 3's, but is similarly symptomatic of an underlying problem with the methodology. The transcript in (13) is taken from the middle of a session with Speaker 4, at a point in time where she had already carried out a number of combination translation/judgment tasks. The reason Speaker 4 gives for preferring the instrumental (=*s*) object is that she believes she uses accusative (=*x̌*) too often (13e, 13k). She goes on to denigrate her own linguistic abilities, saying she 'may forget to speak the language' (13k). She also remarks that she likes the instrumental case 'now that you brought it to my attention' (13k).

(13) Speaker 4 commentary

   a. Speaker: *sa dənəm* ('the rope (INST)').

   b. KS: And is that better than *x̌a dənəm* ('the rope (ACC)').?

   c. Speaker. I like *sa*.

   d. KS: Kay. Does *x̌a*, does it sound weird, or...?

   e. Speaker: No, I just use it too much.

   f. KS: Oh.

   g. Speaker: I like, I'm gonna use *sa*. [...]

   h. KS: [...] Okay. So, so in that case, um – I still wanna know – so for the purpose of teaching...

   i. Speaker: Yeah, sure.

   j. KS: ...that sort of thing, if it's okay to say *x̌a* or *sa*.

   k. Speaker: It's okay. Cuz uh – uh, I may forget to speak the language, and I just use *x̌a* all the time. I like *sa*. Now that you brought it to my attention.

   l. KS: So I wanna know, I wanna know – what's possible, and what you like best.

   m. Speaker: *ʔeʔ* ('Yes.'). Well I like *sa*, cuz I'm not using it, and I want to.
                                                                                              (20160710-S4 VF)

As with Speaker 3, something about the combination translation/judgment task appears to be undermining Speaker 4's confidence in her own ability to make semantic judgments that are

representative of the wider Kwak'wala speech community. Instead of offering her own semantic intuitions, Speaker 4 has shifted to providing beliefs about what might be the correct way of saying things according to a hypothetical Kwak'wala speaker.

3 Explaining the problem

It is important to reiterate that Speaker 3 and Speaker 4 were experienced in carrying out judgment tasks prior to the sessions in which I introduced combination translation/judgment tasks. This suggests that the problems documented in Section 2 are not due to these consultants' inexperience with judgment tasks, but were triggered, rather, by something inherent to the combination translation/judgment task itself, or to the way it was explained to them (more on this point in Section 4). The first mystery to be solved then is the following: what is it about the combination translation/judgment task which led Speaker 3 and Speaker 4 to shift away from offering their own semantic intuitions in judgment tasks?

The first thing to recognize about the combination translation/judgment task is that because of its structure, it is very likely to trigger metalinguistic awareness of what is being studied. This occurs because the constructed sentence is, by design, identical to the consultants' previously volunteered sentence except for one minimal change. In the Kwak'wala situation, whenever the consultant volunteered a sentence with an instrumental object she would subsequently be asked to judge a sentence with an accusative object; and whenever she volunteered a sentence with an accusative object she would subsequently be asked to judge a sentence with an instrumental object. Consultants' attention was immediately drawn to this change in object case marking, leading them to focus their awareness on the difference between instrumental ($=s$) and accusative ($=\check{x}$) marking.[10] This diverted their attention away from the more concrete problem of whether the sentence they were in the process of judging was semantically acceptable. For Speaker 3, this shift towards metalinguistic reflection is apparent in (12) in how she directs her commentary not onto the constructed sentence, but onto her own linguistic abilities. For Speaker 4, this tendency towards abstracted reflection is apparent in her expression of the belief in (13) that in general, she is using accusative case too often.

Yet while metalinguistic awareness of the topic under discussion may lay the groundwork for the issues being examined, it is not the core of the problem. Quite often in elicitation, consultants become at least somewhat aware of what is being researched, but this awareness doesn't necessarily alter the way consultants respond to judgment tasks. Nevertheless, metalinguistic awareness does seem to interact badly with another aspect of the combination translation/judgment task's

---

[10]The case markers $=s$ (instrumental) and $=\check{x}$ (accusative) are often followed by other determiners or strings of determiners, such as $=u\check{x}$ (=D2) or $=u\check{x}da$ (=D2=DET). More precisely then, consultants' awareness became immediately focused on differences between clitic strings encountered in actual examples, such as $=su\check{x}da$ vs. $=\check{x}u\check{x}da$, rather than on the more abstracted difference between $=s$ and $=\check{x}$. I will continue in what follows to refer to the difference between instrumental ($=s$) and accusative ($=\check{x}$), but note that this is for expository convenience as the reality is more complicated than this. In fact, the definition of a 'minimal' change between the volunteered translation and the constructed sentence as outlined in (4) is technically violated in some of my data, as in, for instance, example (6) where the volunteered form has $=\check{x}^wa$ (=ACC=D2=D4) and the constructed sentence has $=su\check{x}$ (=INST=D2). It may be that the difference in determiner semantics between these two clitic strings is insignificant enough in the given context that the consultant was willing to overlook it.

structure, which is that the task involves repeatedly asking consultants to judge sentences that are different from the one they just produced.

Consider that from the elicitor's point of view, the sentence that is put forward to the consultant to be judged is constructed on the fly – that is, its construction is based upon whichever sentence the consultant first volunteered. In my Kwak'wala situation, if the consultant volunteered a sentence with an instrumental object the constructed sentence will contain an accusative object, and if the consultant volunteered a sentence with an accusative object the constructed sentence will contain an instrumental object. From the elicitor's perspective, this relation between a consultant's volunteered sentence and the constructed sentence they will be asked to judge is known and taken for granted. From the consultant's point of view, however, this relation between their own volunteered sentence and the constructed sentence they are asked to judge may not be apparent. The consultant may, upon reflection, arrive at the conclusion that the elicitor is always asking about sentences they themselves did not say because their volunteered sentences are deviant in some way. In other words, the consultant may not realize that this is just the way the combination translation/judgment task is designed, and may read into this task structure the implication that their volunteered sentences are not what the elicitor expects to hear.

Imagine that you are the language consultant. You are repeatedly asked to judge sentences which are minimally different from the sentence you just volunteered. You might wonder: *Why is the researcher always asking about sentences that are different from the one I just said? Are my sentences abnormal?* Keep in mind, also, that as the language consultant you know that the elicitor (here KS) works with other language consultants and is therefore in a position to know whether you speak differently than other consultants. Knowing this, you might wonder: *Is the researcher asking about sentences that are different from mine because that is how she has heard other language consultants speak?*

It is not hard to see how reasoning of this sort could arise naturally in response to the structure of the combination translation/judgment task, where the consultant's volunteered sentence is repeatedly being altered and read back for judgment. Reasoning of this sort then has the potential to undermine the consultant's belief that her judgments are representative of those in the wider speech community. Consider that when Speaker 3 first began doing combination translation/judgment tasks, she was still offering her own intuition in the judgment portion of the task (9). It was only later on in the same session, in the course of doing more combination translation/judgment tasks involving object case substitution, that Speaker 3 shifted to accepting every constructed sentence. By the time the dialogues in (11)-(12) were produced, Speaker 3 had shifted her approach to the judgment task entirely. Not only did she now accept every constructed sentence, but she also commented that each constructed sentence was likely to be better than the translation she had previously volunteered. In other words, she had shifted away from providing her personal semantic intuitions towards providing reflections on what might be possible or preferable within the wider community of speakers. Her way of responding to the judgment portion of the task changed over time as she did more of the combination translation/judgment task.

What might lead some consultants to arrive at the conclusion that their judgments are deviant, while others do not? There are several factors, none of which are mutually exclusive, that could increase the likelihood that a particular consultant will come to this conclusion. Firstly, it is sometimes the case that objectively fluent Kwak'wala speakers will express low confidence in their linguistic abilities, whether in general or on a particular day. A consultant who is feeling

insecure about their fluency is more likely to be susceptible to the implication that their linguistic intuitions are deviant. Secondly, a consultant may be particularly susceptible to the implication that their intuitions are deviant if they are hyperaware that their responses might ultimately be used to teach the language. A consultant in such a state could adopt the observed response pattern out of fear of making errors that could then be passed on to learners. The importance of this factor is suggested by Speaker 3's comment in (11) that a form with instrumental case may be preferable "if you're teaching somebody". Finally, a consultant who is looking for clues about why the elicitor is asking the questions she is might begin to construe the elicitor's wording in a prescriptive light. For instance, in (10) KS uses the phrase "not good" to describe something the speaker wouldn't say (10c), while in (10e) she describes the consultant as an "expert". The possibility that a prescriptive frame has been unintentionally invoked is suggested, for instance, by Speaker 3's comment that what she volunteered is not "the proper sentence" (12). While I have endeavoured for years to instil a descriptive frame in which to study Kwak'wala with my consultants, it is always possible that improper wording choices on my part as elicitor could trigger a shift into a prescriptive frame. Caution should always be taken to avoid wording that could steer consultants away from offering their descriptive intuitions and towards the evaluation of prescriptive norms.[11]

Consultants' reflections on what might be possible within the wider speech community can be interesting in their own right, but they are not a valid source of evidence for semantic hypotheses. This is because in general, it is not possible for speakers of any language to make semantic judgments on behalf of other speakers of the same language. In order for the combination translation/judgment task to be used to generate data that is useful for hypothesis testing, something must be done to make sure that consultants' confidence in their own semantic judgments is not undermined in the course of obtaining these judgments.

4   Preventing the problem

While previously I have identified the problem with the combination translation/judgment task as a structural problem inherent to the task itself, ultimately the task's failure rests on me, the elicitor, for not adequately explaining the task to my consultants.

Successful usage of the combination translation/judgment task relies on the language consultant confidently asserting their native speaker intuitions for the judgment portion of the task. Some examples showing the task being carried out successfully are given in (14)-(15). In (14), the speaker comments that the judged form 'doesn't sound right' (14e) and then reasserts a fragment of the proper form (14g), while in (15), the speaker remarks that the judged form 'sounds legal' (15g). These statements are individual-level judgments expressing these speakers' intuition of what is felicitous in their language. There is no reference in the consultants' commentary below to what other speakers might say about these sentences.

(14)   Speaker 1: Combined translation/judgment task

   a.   KS: And how would you be like, 'Take off that toque!'...?

   b.   Speaker: *lawedas x̌us budzəx̌ƛə{'}ax̌!* [c]

---

[11]I am grateful to my two reviewers for inspiring reflection on the factors discussed here.

c. lawedas　　　　　　　**x̌us**　　　budzəx̌λəⁱax̌!
    lawe　-x?id -a =s **=x̌** =us　budzəx̌λəⁱax̌
    come.off -BEC -A =2 **=ACC** =2.POSS toque
    'Take off your toque (**ACC**)!'　　　　　　　　　　　　　　(20160804-S1 VF)

d. KS: And can you ever say, *lawedas sus budzəx̌λəⁱa...*? [h]

e. Speaker: *laweda?s sus*, no, it doesn't sound right.

f. KS: Kay.

g. Speaker: It's gotta be *x̌us*!

h. #lawedas　　　　　　　**sus**　　　budzəx̌λəⁱax̌
    lawe　-x?id -a =s **=s**　=us　budzəx̌λəⁱax̌
    come.off -BEC -A =2 **=INST** =2.POSS toque
    'Take off your toque (**INST**)!'　　　　　　　　　　　　　(20160804-S1 JF)

(15) Speaker 2: Combined translation/judgment task

a. KS: And if we wanted to specify that it was a green hat...

b. Speaker: Mhm.

c. KS: ...how would we do that? He 'put a green hat on'.

d. Speaker: *λətəmdux̌ Eddie sa łənx̌a λətəmł*. [e]

e. λətəmdux̌　　　　　　　Eddie **sa**　　łənx̌a λətəmł
    λət　-(ǧ)əm -x?id =ux̌ Eddie **=s**　=a　łənx̌
    overhang -face -BEC =D2 Eddie **=INST** =D4　green
    'Eddie put on a green hat (**INST**).'　　　　　　　　　　(20160707-S2 VF)

f. KS: Can I say *λətəmdux̌ Eddie x̌a łənx̌a λətəmł...*? [h]

g. Speaker: *λətəm... λətəmdux̌ Eddie x̌a łənx̌a λətəmł*. [h] That sounds legal.

h. λətəmdux̌　　　　　　　Eddie **x̌a**　　łənx̌a　λətəmł
    λət　-(ǧ)əm -x?id =ux̌ Eddie **=x̌**　=a　łənx̌ -a λətəmł
    overhang -face -BEC =D2 Eddie **=ACC** =D4 green -A hat
    'Eddie put on a green hat (**ACC**).'　　　　　　　　　　(20160707-S2 JF)

    For the combination translation/judgment task to work correctly, clear communication with the consultant is needed on two topics. First, the consultant needs to understand the steps of the task, and second, they should also understand the importance of providing individual intuitions as a native speaker of the language.
    A hypothetical task description which covers these topics is offered in (16).

(16) Sample description: Combination translation/judgment task
    First, I am going to present a situation to you, and then I am going to ask you to translate a sentence from English into Kwak'wala which fits this situation. After that, I am going to take whatever Kwak'wala sentence you just gave me, and I'm going to change it a little bit

and say it back to you. At this point, I would like you to tell me whether or not the changed sentence makes sense in the situation we are talking about. Does the changed version of your sentence still fit the situation we're talking about, or does it sound funny to you? Has the meaning changed? Since I am not a native speaker, I have no way of knowing whether the change I've made to the sentence you gave me makes sense or not. I am interested in your personal judgment of this sentence, as you are a native speaker and possess special knowledge of the language that non-speakers like me do not have.

The sample task description in (16) includes an explanation of how each sentence to be judged is related to a sentence the consultant has volunteered. This information is included because it is important for the consultant to understand that the sentences she will judge are constructed based on the sentences she will volunteer and are not, crucially, sentences that the elicitor secretly believes to be correct. Another point of emphasis in (16) is the importance of providing native speaker intuitions as judgments. This idea is framed in terms of the special knowledge the consultant possesses as a native speaker of the language, the intention being to communicate to the consultant the singular importance of their intuitions. For those who work with multiple consultants, it may also be important to discuss the dimensions of linguistic variation that exist within every language community. Ideally, the task description in (16) should lead into a productive dialogue between the elicitor and the consultant in order to clarify any concepts or parts of the process that are hard to understand.

Explaining the combination translation/judgment task in a clear and informative way provides a mechanism for preventing the problems with the combination translation/judgment task discussed earlier. What about fixing a problem situation that has already arisen, such as in my fieldwork situation? In my case, I decided to stop doing combination translation/judgment tasks with Speaker 3 and Speaker 4, and to date have not reintroduced this method into elicitation with these consultants. While I am optimistic that with greater care and attention paid to the task description, the combination translation/judgment task could be carried out once more with these consultants, I am also wary of the dire consequences if we were to fall into familiar patterns. I have therefore left the question open as to whether it is possible to fix the problem once it has arisen.

Finally, it is worth noting that in order to avoid the perils of the combination translation/judgment task entirely, one could achieve the same outcome by eliciting translations in one session and the associated judgments in a different session. In my own fieldwork, I have often done this. The downside of this bifurcation is that it takes almost twice as long to elicit the same data, since it requires that each context be introduced in two different sessions. This downside becomes significant when there are limitations to the time or resources available to engage in fieldwork, as may often be the case when working with elderly speakers of endangered languages.

5   Conclusion

While the implementation of the combination translation/judgment task described in this paper, as well as the problems this implementation led to, were specific to my Kwak'wala fieldwork situation, the methodological issues raised here are relevant for any fieldworker interested in adapting this method to their fieldwork situation. When carried out successfully, the combination translation/judgment task can be extremely useful for generating ample semantic data in a relatively

short amount of time. However, the discovery that the structure of the task can bias even experienced language consultants towards questioning their own semantic judgments means that great care must be taken to execute the task correctly.

We have seen that the combination translation/judgment task has a high likelihood of generating metalinguistic awareness of the phenomenon being studied. While this is something to be aware of when using this methodology, it is not a reason to discount it. For two of my language consultants using this method, having heightened awareness of object case marking did not affect the way they carried out the task. For two other consultants, however, heightened awareness of object case marking interacted with the structure of the task, ultimately leading to the problems I've described above. To mitigate the risks associated with generating metalinguistic awareness, the elicitor should aim to keep the amount of time spent in metalinguistic reflection during a session to a minimum. To this end, the combination translation/judgment task should be carried out alongside other methods of data collection which have less of a propensity to generate metalinguistic awareness, such as translation tasks in isolation from judgment tasks or storyboards. Moreover, because it is iteration of the combination translation/judgment task which can lead to trouble, care should always be taken to avoid immoderate iteration of the task in any single session.

Generalizing beyond the fieldwork context explored here, this paper should serve as a cautionary tale for all fieldworkers about the perils of repeatedly asking consultants to judge sentences that are minimally different from the sentence the consultant just volunteered. I have argued that iterated judgments of this sort carry the risk of implying that the consultants' volunteered sentences are deviant in some way. This, in turn, can lead some consultants to lose confidence in their semantic judgments – or, at least, in their judgments being representative of the wider speech community. The result is that consultants end up shifting away from offering their semantic intuitions in the judgment portion of the task, towards offering guesses about what might be semantically acceptable within the wider community. Not only does this process result in unusable data for the researcher, but it can result in much worse. Fieldworkers have an ethical responsibility to support and empower language consultants in sharing their knowledge and expertise. If the combination translation/judgment task leads to a situation that undermines a consultant's confidence in their linguistic abilities, then there is no worse possible outcome from an ethical perspective.

Fortunately, by being careful and informative in the way we explain the task to consultants we can prevent this worst possible outcome. We may also benefit from seeing the process of describing tasks to our consultants in a new way – as an opportunity to communicate our appreciation for the invaluable intuitive knowledge that language consultants are willing to share with us.

References

Bohnemeyer, Jürgen. 2015. A practical epistemology for semantic elicitation in the field and elsewhere. In *Methodologies in semantic fieldwork*, ed. by Ryan Bochnak and Lisa Matthewson, 13-46. Oxford: Oxford University Press.

Matthewson, Lisa. 2004. On the methodology of semantic fieldwork. *International Journal of American Linguistics* 70(4). 369-415.

Sardinha, Katie. 2016. The semantics of object case marking in Kwak'wala. In *SULA 9: Pro-*

ceedings of the Ninth Conference on the Semantics of Under-represented Languages in the Americas, ed. by Thuy Bui and Rudmila-Rodica Ivan, 155-172. Amherst, MA: University of Massachusetts, Graduate Linguistics Student Association.

Sardinha, Katie. 2017. The semantics of Kwak'wala object case. Doctoral dissertation, University of California, Berkeley.

Sardinha, Katie. 2018. Viewpoint aspect and object case in Kwak'wala and Finnish. In *Papers for the 44th Meeting of the Berkeley Linguistics Society*, ed. by Garvin, Karee, Noah Hermalin, Myriam Lapierre, and Yevgeniy Melguy, 271-285. Berkeley: University of California, Berkeley Linguistics Society.