Subject obviation and self-locating knowledge

Julie Goncharov

COCOA, March 08, 2024

Nay, if thy wits run the wild goose chase, I have done;
for thou hast more of the wild goose in one of thy wits than,
I am sure, I have in my whole five...
W. Shakespeare *Romeo and Juliet*, Act 2 scene 4

,,

1 Background

(1) a. #Je veux que je parte. 'I want that I leave-SBJV' *cf.* Je veux partir. (Ruwet 1991)
     b. Je veux absolument que j'amuse ces enfants. 'I want absolutely that I amuse-SBJV these children'

(2) a. #Penso che io parta domani. 'I think that I leave-SBJV tomorrow' (Costantini 2016)
     b. Penso che io abbia fatto molti errori. 'I think that I have-SBJV made many mistakes'

(3) a. #Remélem, hogy (nem) szédülök 'I hope that I (don't) have-IND vertigo' (Szabolcsi 2021)
     b. Remélem, hogy nem untatlak. 'I hope that I'm not boring-IND you'

It has been already established in the literature[1] that subject obviation

- is not due to some syntactic restriction (because of the amelioration effects)
- is not due to a competition between the subjunctive and the indicative mood (because it is not restricted to subjunctive clauses)
- is not due to a competition between finite clauses and infinitives (because for some cases, there are no infinitival alternatives)
- is not restricted to decisive modality, such as, desires, hopes, commands (because subject obviation is found with epistemic attitudes)

Two promising insights about the nature of subject obviation:

1. Ruwet (1991), Schlenker (2011), Costantini (2016), Szabolcsi (2021)
   subject obviation is somehow related to the impossibility of "distancing oneself", "event de se interpretation", "introspective knowledge", or "direct experience"

2. Kaufmann (2019)
   subject obviation is due to a conflict between two pragmatic requirements on the Common Ground; in 'α attitude that $F(\alpha)$'

   - on the one hand, α is certain whether $F(\alpha)$
   - on the other hand, the attitude requires that α is uncertain whether $F(\alpha)$

---

[1] See Schlenker (2005), Schlenker (2011), Costantini (2016), Szabolcsi (2021), Goncharov (2023).

2   In this talk

Part 1: Subject obviation and essentially private information

(4)   a. Remélem, hogy benne vagyok a csapatban.   'I hope that I'm on the team'
        b. Remélem, hogy nem untatlak.   'I hope that I'm not boring you'
        c. Remélem, hogy biztonságban vagyok.   'I hope that I'm safe'
        d. Remélem, hogy egyenesen állok.   'I hope that I'm standing straight'
        e. #Remélem, hogy fél lábon állok.   'I hope that I'm standing on one leg'
        f. #Remélem, hogy (nem) szédülök.   'I hope that I (don't) have vertigo'
        g. #Remélem, hogy (nem) fázom.   'I hope that I'm (not) cold'
        h. #Remélem, hogy ugrándozok.   'I hope that I'm jumping around'
        i. #Remélem, hogy próbálok segíteni.   'I hope that I'm trying to help'
        j. #Remélem, hogy simogatom a macskát.   'I hope that I'm stroking the cat'
        (Szabolcsi 2021, 10)

(5) A and B are talking in plain view of each other. B says to A:

        a. I have vertigo.
        b. #I am jumping around.

(6) I believe I am sane.

Certainty and uncertainty

(7)   a. I think I am boring you.
        b. I think I am safe.
        c. #I think I am jumping on one leg.
        d. #I think I am petting the cat.

Part 2: Subject obviation and intentions

(8)   a. #Je veux que je parte.   'I want that I leave'
        b. Je veux que je guérisse.   'I want that I get-better'
        c. Je veux que je réussisse.   'I want that I succeed'
        (Ruwet 1991, 25-26)

(9) Said by the lady of the house:
      #Je veux que je reçoive mes invités avec faste.
      'I want that I receive my guests with pomp'   (ibid., 25)

(10) Said by Napoleon to an aide-de-camp he is sending to Davout, who commands another detachment:
      Dites-lui bien que je veux que je reçoive son messsage dans les plus brefs délais.
      'Tell him that I want that I receive his message without delay'   (ibid., 25)

3   A quick look at self-locating knowledge

(11) Rudolf Lingens, a famous Stanford amnesiac, who has read a complete biography of himself and among other things, learned that Rudolf Lingens was born in 1953, says or thinks to himself:
      a. I know that Lingens was born in 1953.
      b. #I know that I was born in 1953.

Lewis's system[2] in which self-located knowledge is captured in terms of centered possible worlds (or properties, broadly construed) succeeds in modelling:

- objective knowledge
- self-locating knowledge
- the relation between objective and self-locating knowledge
- integration of one person's self-locating knowledge with objective knowledge of the same person

... but Lewis's system is solipsistic and does not allow to model integration of objective and self-locating knowledge of different individuals (Stalnaker 2008, 2014)

(12) Daniels who knows that the famous Stanford amnesiac is Rudolf Lingens and also knows that Lingens was born in 1953, comes to Lingens and says: "You were born in 1953". Now, Lingens can sincerely say or think to himself:
   a. I know that I was born in 1953.

Stalnaker's modifications

A (partial) model (ignoring times) is a tuple $\langle W, S, E, R \rangle$ where

- W is a set of possible worlds (possibilities, properties of the Universe)
- S is a set of subjects or agents
- E is a set of centered worlds meeting the following condition: A of c exists in w of c
- R is a binary relation on E (transitive, Euclidean, serial)

A centered world is a pair $\langle c, w \rangle$ where $c \in S$ and $w \in W$.

Modification 1: subject specification in the relata

The accessibility relation (only one) is established between centered worlds, e.g., $\langle A, x \rangle R \langle B, y \rangle$ means that A in x locates herself as B in y.

NB: Unlike for Hintikka and Lewis, for Stalnaker the subject specification is in the relata and not on the accessibility relation. This modification addresses the second issue above.

Hintikka: y is compatible with what A knows in x iff $xR_A y$

Stalnaker: y is compatible with what A knows in x iff for some center c, $\langle A, x \rangle R \langle c, y \rangle$

A belief state is defined as usual as a set of centered worlds R-related to the base centered world.

Modification 2: different centers - different worlds

for all $c, c', c'' \in S$ and all $x, y \in W$, if $\langle c, x \rangle R \langle c', y \rangle$ and $\langle c, x \rangle R \langle c'', y \rangle$, then $c' = c''$

This assumption ensures that if a center locates herself differently, there are two different possibilities. In other words, there is a one-one correspondence between centered possible worlds and uncentered possible worlds.

NB: This is different from Lewis's view, for whom centered possible worlds are fine-grained representations of the same possible world. This modification allows to avoid the first issue above.

Modification 3: Common Ground includes self-locating information

Common Ground, which is usually defined as common knowledge (i.e., an iterative attitude A knows that B knows that A knows...), is modelled as a set of multi-centered possible worlds where one center represents

[2]Lewis (1979)

who A takes herself to be in the wolds according to B, another center for representing who A thinks B takes himself to be in those worlds, etc.

(13) Rudolf Lingens, a famous Stanford amnesiac, who has read a complete biography of himself and among other things, learned that Rudolf Lingens was born in 1953, says or thinks to himself:
   a. I know that Lingens was born in 1953.
   b. #I know that I was born in 1953.

f = an 'I-concept', i.e., an individual concept (a function from W to S)
C = a set of uncentered possible worlds corresponding to the set of multi-centered possible worlds that constitute the Common Ground
L = Lingens
O = O'Leary w = the actual world

(14)   a. I know that Lingens was born in 1953.
   b. $\forall x \in C, c \in S : \langle L, w \rangle R \langle c, x \rangle \rightarrow x \in \phi$
   c. $\phi = \{ x : \text{Lingens is born in 1953 in } x \}$
   d. $\langle \langle L, w \rangle, \langle L, w' \rangle \rangle, \langle \langle L, w \rangle, \langle O, w'' \rangle \rangle \in R$
   e. $f(w') = L, f(w'') = O$

(15)   a. #I know that I was born in 1953.
   b. $\forall x \in C : \langle L, w \rangle R \langle f(x), x \rangle \rightarrow x \in \phi$ $w'' \notin \phi$
   c. $\phi = \{ x : f(x) \text{ is born in 1953 in } x \}$
   d. $\langle \langle L, w \rangle, \langle L, w' \rangle \rangle, \langle \langle L, w \rangle, \langle O, w'' \rangle \rangle \in R$
   e. $f(w') = L, f(w'') = O$

(16) Daniels who knows that the famous Stanford amnesiac is Rudolf Lingens and also knows that Lingens was born in 1953, comes to Lingens and says: "You were born in 1953". Now, Lingens can sincerely say or think to himself:

   a. I know that I was born in 1953.

D = Daniels

(17) Initial CG:
   a. Daniels knows that Lingens thinks to himself "I was born in 1953 if I am Lingens and I was born in 1954 if I am O'Leary."
   b. $\exists x \in C : \langle (D, L), w \rangle R \langle (D, f(x)), x \rangle \land x \in \{ y : f(y) \text{ is born in 1953 in } y \}$
   c. $\exists x \in C : \langle (D, L), w \rangle R \langle (D, f(x)), x \rangle \land x \in \{ y : f(y) \text{ is born in 1954 in } y \}$
   d. $f(w') = L, f(w'') = O$
(18) Daniels to Lingens:
   a. You were born in 1953.
   b. $\phi = \{ x : g(x) \text{ is born in 1953 in } x \}$ $\forall x \in C : g(x) = f(x)$
(19) Updated CG:
   a. Daniels knows that Lingens (now) thinks to himself "Oh, I was born in 1953."
   b. $\forall x \in C : \langle (D, L), w \rangle R \langle (D, f(x)), x \rangle \rightarrow x \in \{ y : f(y) \text{ is born in 1953 in } y \}$
(20) Lingens to himself:
   a. I know that I was born in 1953.
   b. $\forall x \in C : \langle L, w \rangle R \langle f(x), x \rangle \rightarrow x \in \psi$
   c. $\psi = \{ x : f(x) \text{ is born in 1953 in } x \}$
   d. $f(w') = L$

4 Subject obviation and essentially private information

Re-formulating Kaufmann's (2019) insight

- **Uncertainty:** because every instance of subject obviation involves an ascription of self-locating information, it is necessary that the belief state of the believer contains worlds in which the ascribed proposition is true as well as worlds in which the ascribed proposition is false

(21) I believe I am sane.

Stalnaker 2008, 65: "self-locating information can exclude certain possibilities only if the subject locates herself in the situation that is excluded"

(22) A and B are talking in plain view of each other. B says to A:
    a. I have vertigo.
    b. #I am jumping around.

- **Certainty:** with essentially private information, e.g., proprioception, there is a strong tendency to assume that the holder of this information is certain about her own internal experiences

(23) In non-pathological, non-dynamic cases, in which A is fully aware of her perceptions and decisions, A thinks to herself:
    a. I think I am boring you.
    b. I think I am safe.
    c. #I think I am jumping on one leg.
    d. #I think I am petting the cat.

Thus, self-ascribing this "privileged" information creates a conflict.

More formally. . .

(24) Uncertainty: A's initial belief state for a successful ascription of self-locating information

    a. A thinks to herself "I am F if I am A1 and I am not-F if I am A2".
    b. $\exists x \in C : \langle A, w \rangle R \langle f(x), x \rangle \land x \in \phi$
    c. $\exists x \in C : \langle A, w \rangle R \langle f(x), x \rangle \land x \notin \phi$
    d. $\phi = \{ x : f(x)$ is jumping on one leg/is petting the cat in $x \}$
    e. $\langle \langle A, w \rangle, \langle A1, w' \rangle \rangle, \langle A, w \rangle, \langle A2, w'' \rangle \rangle \in R$
    f. $f(w') = A1, f(w'') = A2$

(25) Certainty: A is the "privileged" source of essentially private information

    a. A thinks to herself "I am F".
    b. $\forall x \in C : \langle A, w \rangle R \langle f(x), x \rangle \to x \in \phi$
    c. $\phi = \{ x : f(x)$ is jumping on one leg/is petting the cat in $x \}$
    d. $\langle \langle A, w \rangle, \langle A, w' \rangle \rangle, \langle A, w \rangle, \langle A, w'' \rangle \rangle \in R$
    e. $f(w') = A, f(w'') = A$

(26) $\forall x \in C : \langle (B, A), w \rangle R \langle (B, f_1(x)), x \rangle \leftrightarrow \langle (B, A), w \rangle R \langle (B, f_2(x)), x \rangle$
    where $f_1 \neq f_2$

(27) Uncertainty: A's initial belief state for a successful belief ascription of non-self-locating information

    a. A thinks to herself "I am F or I am not-F".
    b. $\exists x \in C, \forall c \in S : \langle A, w \rangle R \langle c, x \rangle \land x \in \phi$

    c. $\exists x \in C, \forall c \in S : \langle A, w \rangle R \langle c, x \rangle \wedge x \notin \phi$
    d. $\phi = \{\, x : f(x)$ is not boring you/is safe in $x \,\}$

(28) Uncertainty: A's initial belief state for a successful belief ascription of objective information for non-obviative sentences

    a. A thinks to herself "B is F or B is not-F".
    b. $\exists x \in C, \forall c \in S : \langle A, w \rangle R \langle c, x \rangle \wedge x \in \phi$
    c. $\exists x \in C, \forall c \in S : \langle A, w \rangle R \langle c, x \rangle \wedge x \notin \phi$
    d. $\phi = \{\, x : $ B is F in $x \,\}$

5 Subject obviation and intentions

Anscombe 1957: "expression of intention" vs. "estimation of the future"

> "if I say"I am going to fail in this exam." and someone says "Surely you aren't as bad at the subject as that". I may make my meaning clear by explaining that I was expressing an intention, not giving an estimate of my chances."

Strong cognitivism (Marisic and Schwenkler 2018, Schwenkler 2021):

> "one intends to do something if and only if one believes in light of one's *practical reasoning* that one will do it. By contrast, one foresees, or predicts that one will do something only if one believes in light of *evidence* that one will bring it about, that one will do so." (Marusic and Schwenkler 2018: 329)

> "An agent's practical beliefs thus comprise a causal order, and the form of causality is that of instrumental teleology: One who believes that she is going to ϕ (drive to the theatre, say) because she is going to ψ (say, see a movie) understands her ϕ-ing not as something that will be *produced* by her ψ-ing, but rather as a suitable *means* to it. And we identify this instrumental order in the agent's view of her future with the content of her intentions." (Marusic and Schwenkler 2018: 329-330)

'Naive' rationalization and self-locating believes (Schwenkler 2024):

(29) I am flipping the switch because I am turning on the light.

. . . but if we know that the light is broken

(30)   a. #She is flipping the switch because she is turning on the light.
    b. She is flipping the switch because she intends (wants) to turn on the light.
    c. She is flipping the switch because she thinks it will turn on the light.

Similarly,

(31) I am packing a bag because I am taking a trip tomorrow.

. . . but if we know that the trip has been canceled

(32)   a. #She is packing a bag because she is taking a trip tomorrow.
    b. She is packing a bag because she plans to take a trip tomorrow.

6 Avoiding a wild goose chase

Kaufmann (2019):
subject obviation is due to a conflict between two pragmatic requirements on the Common Ground;
in 'α attitude/imp that $F(\alpha)$'

- 'α is certain whether $F(\alpha)$' is the product of Decisive Modality and Control conditions
- 'α is uncertain whether $F(\alpha)$' is the condition that comes from 'attitude/imp'

. . . but decisive modality is not necessary

(33)   a. #Penso che io parta domani. 'I think that I leave-SBJV tomorrow'
        b. Penso che io abbia fatto molti errori. 'I think that I have-SBJV made many mistakes'

. . . nor is attitude ascription itself

(34) A and B are talking in plain view of each other. B says to A:
        a. I have vertigo.
        b. #I am jumping around.

Costantini (2016) and Szabolcsi (2021):
factives also show subject obviation, but they cannot be explained as a conflict between two pragmatic requirements on CG (because CG entails the content of the attitude ascription); move to counterfactual reasoning

(35)   a. #Sajnálom, hogy ugrándozok. 'I regret that I am-IND jumping around'
        b. Sajnálom, hogy untatlak. 'I regret that I am-IND boring you'

Changing the question

- What makes particular modalities or moods to result in subject obviation?
- What prevents particular modalities or mood from allowing the uncertainty accommodation?

References

Costantini, Francesco. 2016. "Subject Obviation as a Semantic Failure: A Preliminary Account." *Annali Di Ca' Foscari. Serie Occidentale* 50: 109–31. https://doi.org/10.14277/2499-1562/AnnOc-50-16-5.

Goncharov, Julie. 2023. "Decisive Modality and Intentionality Effect." *Canadian Journal of Linguistics*.

Kaufmann, Magdalena. 2019. "Who Controls Who (or What)." In *Proceedings of SALT 29*, edited by Katherine Blake, Forrest Davis, Kaelyn Lamp, and Joseph Rhyne, 636–64. https://doi.org/https://doi.org/10.3765/salt.v29i0.4643.

Lewis, David. 1979. "Attitudes de Dicto and de Se." *The Philosophical Review* 88 (4): 513–43. https://www.jstor.org/stable/2184843.

Ruwet, Nicolas. 1991. *Syntax and Human Experience*. Chicago: University of Chicago Press.

Schlenker, Philippe. 2005. "The Lazy Frenchman's Approach to the Subjunctive." In *Romance Languages and Linguistic Theory*, 269–309.

———. 2011. "Indexicality and de Se Reports." In *Semantics: An International Handbook of Natural Language Meaning*, edited by Claudia Maienborn, Klaus von Heusinger, and Paul Portner, 1561–1604. Berlin: Mouton de Gruyter.

Stalnaker, Robert C. 2008. *Our Knowledge of the Internal World*. Oxford: Oxford University Press.

———. 2014. *Context*. Oxford: Oxford University Press.

Szabolcsi, Anna. 2021. "Obviation in Hungarian: What Is Its Scope, and Is It Due to Competition?" *Glossa: A Journal of General Linguistics* 6 (1): 1–28. https://doi.org/10.5334/gjgl.1421.