Face Effects Annotation Manual
(Short version)

2025, S. Amido and A. Soubki

Introduction

This document serves as a manual for an annotation of face effects based on the theory of politeness presented by Brown & Levinson (1987). Their framework attempts to provide a way of reasoning about the actions of agents primarily through the lens of **face** which they define as,

> *The public self-image that every member [of a society] wants to claim for himself* (Brown & Levinson, 1987, §3.1, p. 61)

They observe that, across cultures, verbal exchanges are not only understood by the explicit meaning of the utterances. Strategies which mitigate impositions or soften offenses, appear constant across cultures, pointing to some universal process (i.e. politeness) being used by agents to produce them.

Brown & Levinson (1987) provide a formal model which accounts for this social calculus. The goal of the following sections is to develop an annotation which can capture the details necessary to (1) study this calculus in action and (2) build computational systems which are able to recreate such universal behaviors seen in natural conversation.

Face and face effects

In this section, we provide the background relevant to the annotation system, which is based on Brown & Levinson's (1987) politeness theory, but includes important distinctions.

Face

Brown & Levinson describe face as "the public self-image that every member [of a society] wants to claim for himself" with two related aspects.

(1) Negative face - *The basic claim to territories, personal preserves, rights to non-distraction - i.e. to freedom of action and freedom from imposition.*
(2) Positive face - *The positive consistent self-image or 'personality' (crucially including the desire that this self-image be appreciated and approved of) claimed by interactants.*

They also provide a definition of face in terms of wants.

(1) Negative face - *The want of every 'competent adult member' that his actions be unimpeded by others.*
(2) Positive face - *The want of every member that his wants be desirable to at least some others.*

To summarize, one aspect (positive face) has to do with recognition, while the other (negative face) deals with obligations and freedoms.

We adopt and build on Brown & Levinson's definition of face and we use more intuitive names for its two aspects. **Connection face** (fka positive face) is understood roughly as one's public self-image, approved of by and desirable to others, and **autonomy face** (fka negative face) is understood as one's autonomy and freedom from imposition by others.

Imagine that each person has some sort of a multi-dimensional (but malleable) shape in their head which represents how they wish others to perceive them. This shape is their face and it is hidden from others. As people interact with the world, they perform actions which others will use to build their own version of what this hidden shape looks like. That is, a speaker will have a shape of themselves in their head as well as a shape for each of the hearers. Likewise, the other dialogue participants will also construct in their heads a shape associated with themselves and with others.

Intentions vs. effects

According to Brown & Levinson, a *face act* is an act, i.e. that which is intended to be done by some verbal or non-verbal communication, which inherently interacts with the face of the speaker and/or hearers (Brown & Levinson, 1987, §3.2, p. 65). According to the authors, a discourse action can threaten the (connection or autonomy) face of the speaker or hearer. Since face acts are defined by their speaker's intention, then it follows that, according to Brown & Levinson, the intention directly interacts with face.

We find that this way of thinking does not capture interpersonal dynamics which condition what is said in interaction since there are cases when the intention that a speaker has is virtually irrelevant. What matters, regarding what happens to face and thus how interaction between interlocutors unfolds, is how the hearer interprets the discourse act. Sometimes, the intention behind the

discourse action and the face effect match up, but sometimes they do not. This is why we are NOT annotating speaker intentions, but rather the effect that the discourse acts have on the hearer.

To illustrate, imagine two interlocutors, Ana and Bob. They are colleagues and have scheduled a meeting in Ana's office at 2pm. When Bob shows up 5 minutes early, Ana utters "Wow, you're early!" Her intention might actually be to compliment Bob on his punctuality, but Bob replies with "Sorry, I didn't mean to intrude. I'll come back at 2pm." Bob didn't interpret Ana's utterance as a compliment, but rather as a complaint. What should be annotated for Ana's utterance is the actual effect it had, which anchors on how it was received by Bob. The face effect(s) associated with a complaint is annotated, and not the face effect(s) associated with a compliment.

Thus, speaker intentions are only important to the extent that they are reconstructed by the hearer. How they are reconstructed is one of the things the annotator considers when making inferences about how interlocutors interpret discourse acts. When annotating the face effects for Ana's utterance, what is relevant is not Ana's actual intention but rather the intention Bob ascribed to her in that moment. The annotator infers Bob's interpretation based on his reaction. In sum, interlocutor intentions are informative for the annotator, but what is annotated are the face effects of each turn (see section Intended vs. actual face effects for more detail on this distinction).

Face effects

A **face effect** is the actual effect (regardless of intention) that an action has on the face of the speaker and/or hearers. An utterance (or other communicative action) can therefore enhance (+) or damage (-) the connection (Con) or autonomy (Aut) face of the speaker (S) or hearer (H), resulting in eight possible face effects. This short version of the manual only considers the four face-*damaging* (-) effects. Face effects are part of the perlocutionary force of speech acts.

To illustrate, imagine our two interlocutors, Ana and Bob. If Ana performs a (discourse) action which causes Bob to construct a shape which is different from what Ana would likely want, this damages her face. Bob cannot see into Ana's head and therefore does not know for sure how Ana wants to be perceived. Neither is it the case that the annotator can see into Bob's head and hence know what shape Bob holds of Ana. However, inferences can be made based on how Bob reacts to Ana's action.

Below are the four relevant face effects, followed by their description, as applied to the situation where Ana is the speaker that performs the action (i.e. the utterance), and Bob is the hearer. Therefore, the effects are shaped in Bob's mental state.

Connection face effects

Connection face is associated with the desire to be approved of by others.
[SCon-] Speaker connection face damaging If Ana's action causes Bob to think of her differently than she would likely wish, this damages Ana's connection face.
[HCon-] Hearer connection face damaging If Ana's actions cause Bob to believe that another (possibly Ana) thinks of him differently than what Bob would likely wish, this damages Bob's connection face.

Autonomy face effects

The state of believing (to some degree) that one is obligated or able to do something is called negative face.
[SAut-] Speaker autonomy face damaging If Ana's actions cause Bob to think of her as more obligated to act in a particular way (different from what Ana would likely wish), this damages Ana's autonomy face.
[HAut-] Hearer autonomy face damaging If Ana's actions cause Bob to believe that he is more obligated to act in a particular way (different from what Bob would likely wish), this damages Bob's autonomy face.

As for the definitions of autonomy face damaging effects, "more obligated" means that one is more obligated than the time prior to the action.

The annotator's expectations

As we have established, neither the speaker nor the annotator can access the hearer's cognitive state. So for any given turn, the annotators must make inferences about how it was received by hearers based on their reactions. By *received* we mean how the hearer interpreted the utterance, and how that interpretation shaped the faces in the hearer's mental state.

In order to illustrate, we can look at how effects on speaker connection face work. Speaker connection face damage is defined as *A discourse action which causes a hearer (not necessarily the addressee) to think of the speaker differently than the speaker would likely wish*. So speaker connection face effects are actually a projection of what the hearer believes happened to the speaker's own connection face in the speaker's head.

Picture once more Ana and Bob, who are work colleagues. They are talking about their summer plans. Ana says she plans to quit her job and go to India to take a yoga teacher training course, a plan that is somewhat popular among her demographic. Now consider the following alternative reactions from Bob and the speaker connection face effects they would have in each case.

Reaction 1
Bob: [sarcastically] “Great! If there’s one thing the world needs, it’s more yoga teachers…”

Bob has judged Ana in a negative light, which (the annotator assumes) Bob assumes Ana does not want, so Ana's connection face is damaged (**SCon-**) as a result of her utterance.

Reaction 2
Bob: “Wow, that’s so cool! I wish I had the guts to make a bold decision like that.”

In this case, Bob sees Ana’s initiative as something admirable, which presumably Ana wants, so Ana’s connection face is not damaged.

Reaction 3
Bob: “Where in India?”

Here, there is no direct appraisal for what Ana has said. In this case, the annotator must infer the effect on Ana’s face within Bob’s mental state based on other available contextual information (i.e., previous and subsequent discourse, multimodal information, Bob’s beliefs and attitudes toward similar things, the relationship between the interlocutors, etc.).

These examples show that the expectations of the annotator play an important role in inferring what face effects utterances have, given that the interlocutors' mental states are not accessible, specifically the hearer's. Note that this is parallel to how expectations regarding other interlocutors (their beliefs and wants) condition and guide our participation in an interaction, in view of what we want to achieve in the conversation. If the annotator expects an utterance to be embarrassing through the hearer's perspective, given what they know about the situational context and the hearer, and if there is no evidence against this expectation, then the annotator will follow through with that judgement. **The annotator's job is, therefore, to make informed guesses about both how the hearer wants to be perceived, and how the hearer imagines the speaker wants to be perceived.**

Representation

To properly annotate a face effect requires the following pieces of information.

(1) Speaker `[str]`
- The speaker of the utterance that caused the face effect.

(2) Source Utterance `[str]`
- The utterance which initiated the face effect.

(3) Hearer `[str]`
- The hearer of the utterance (the addressee or a bystander), whose mental state is being affected by the face effect.

(4) Hearer / Speaker Orientation `[bool]`
- Is the effect damaging the face of the hearer or speaker of the utterance?

(5) Connection / Autonomy Face Orientation `[bool]`
- Is the effect damaging connection or autonomy face?

We model a face effect as a 5-tuple which is uniquely described by the information above. The speaker, hearer, and source utterance (1,2, and 3 respectively) will typically be provided to the annotator. This leaves the hearer/speaker and connection/autonomy face orientations (4, 5) to be completed by the annotators. A single source utterance may have several face effect annotations (i.e., it may appear in multiple 5-tuples).

While we have summarized the differences between these orientations which require annotation in the sections above, the following sections will provide more specific guidelines for identifying the presence of the four face damaging effects.

Annotation guidelines

In this section we detail guidelines for the annotation of the four face damaging effects. For each face effect we will provide (1) a broad definition describing how face is affected (2) criteria in the form of specific effects or related intended actions, and (3) some concrete examples from data.

In particular, we draw examples from the Wikipedia Talk Pages Corpus (WikiTalks) and the TOWIE corpus (transcribed dialogues from The Only Way is Essex, a British reality television show).

It is important to note that the examples do not necessarily mean that all instances of the described discourse action should receive that face effect label. In other words, there may be exceptions to the examples provided. Furthermore, examples below may have other face effect labels in addition to the face effect being described.

Intended vs. actual face effects

We want to annotate the *actual* face effect, not the *intended* face effect of an utterance. This requires looking at the subsequent turn to determine the effect of both hearer and speaker orientated face effects. The example below illustrates the crucial difference between intended and actual face effects.

Mean Girls

Speaker	**Utterance**	**Hearer**	**Face Effect(s)**
Regina	But you're, like, really pretty.	Cady	HAut-
Cady	Thank you.	Regina	SAut-, SCon-
Regina	So you agree?	Cady	HCon-, HAut-
Cady	What?	Regina	SCon-
Regina	You think you're really pretty?	Cady	*HCon-, HAut-*[1]

In this example from *Mean Girls* (M. Waters, 2024), Regina's plan is to intimidate Cady by taking advantage of the fact that compliments normally elicit two socially conflicting responses. On the one hand, one can respond to a compliment by thanking the compliment-giver. However this can

[1] Note that since face effects are determined based on the reaction to the source utterance, then the last turn in a dialogue cannot technically be annotated, given the lack of reacting turn. However, given that this is an excerpt of a larger dialogue that is accessible, the grey italic font indicates that it is based on the (not shown) subsequent turn.

be interpreted as an implicit agreement with the compliment, which in turn conveys a lack of modesty. On the other hand, one can respond to a compliment by rejecting it, and thus coming across as modest. However, rejecting the compliment can also show ingratitude. Therefore, there are two socially desirable qualities in competition: a display of modesty versus a display of gratitude.

Regina baits Cady with a compliment so that whichever response Cady provides (thanking or rejecting), Regina can call attention to either her lack of modesty or her ungratefulness. Following Cady's thanking, Regina challenges her alleged lack of modesty.

We evaluate the effect of the compliment by looking at the subsequent line, where Cady thanks Regina for the compliment. So while the intention of Regina was all along to bully Cady (HCon-) as evidenced by the subsequent turns, there is no such effect because at this point Cady is unaware of Regina's multi-turn strategy.

Regina's challenge in line 3 is annotated as HCon- because it causes embarrassment for Cady, and HAut- because it makes her more obligated to explicitly and publicly commit to a self-compliment, an undesired situation since it displays lack of modesty.

Below, in section Face effects: definitions, criteria, examples, definitions of face effects are followed by *Criteria*. Criteria include two things: actions which usually, but not always, lead to the face effect in question, and descriptions of effects which fall within the larger face effect in question. For example, the reader will not find the action of complimenting under the hearer connection face damaging (HCon-) description because this is not an effect that actions intended as compliments usually have. However, to reiterate, effect annotation is anchored to evidence in the reaction. Therefore, if the action intended as a compliment is instead received as an insult, then it must be annotated as such, i.e., as hearer connection face damaging (HCon-).

Face effects: definitions, criteria, examples

[SCon-] Speaker connection face damaging

Definition

A discourse action which causes H (not necessarily the addressee) to think of S differently than S would likely wish.

Note that this is often involuntary or unintended.

Criteria[2]

1. *Those actions that are damaging to S's face because of their basic positive-face wants of self-control and self-respect (Brown & Levinson, 1987, p. 286; fn 13)*
    (a) *apologies* (S allows H to see them as someone who has made a mistake or transgression or reveals some undesirable quality – see also SCon+ criteria 1d)
    (b) *breakdown of physical control over body, bodily leakage, stumbling or falling down, etc.*; *emotion leakage, non-control of laughter or tears*
    (c) *self-humiliation,* mistakes*, shuffling or cowering, acting stupid, self-contradicting*, bragging, and other forms of embarrassing behavior (S allows H to recognize undesirable qualities in S) – "Thanks, my pies are pretty good" (bragging) (S shows a, possibly socially inappropriate, lack of modesty to H)
    (d) confessions, admissions of guilt or responsibility - e.g. for having done or not done an act, or for ignorance of something that S is expected to know
2. S attempts to improve H's view of S but fails to do so.
    (a) poorly received actions associated with an intended HCon+ such as compliments, agreement, understanding, apologies, etc. which are met with offense – "You're pretty good at programming for a girl. - What's that supposed to mean?" (S is seen undesirably by H due to their misunderstanding of H's connection face wants – This may be accompanied by HCon-.)
    (b) apparent insincerity in actions (e.g. apologies, compliments, offers, etc. – perhaps accompanied by stark exaggeration – "Thank you so very much for bringing these twinkies to our soiree.", bragging, name dropping, etc. (S attempts to improve H's view of S but fails to do so)

Example

WikiTalk

Speaker	Utterance	Hearer	Face Effect(s)
Kelly	Jesus, calm down, Jossi.	Jossi	HAut-, HCon-
Kelly	I admit to stalking Rootology's contribs since he always edits interesting articles.	Jossi	**SCon-**

[2] In the Criteria section of all four face damaging effects, the majority of the content is from Brown & Levinson (1987, pp. 65-67) with minor adaptations. Content by Brown & Levinson (including adapted gender-neutral language, term replacement of *acts* with *actions*, and relevant omissions) is italicized. Our added criteria and/or examples are not italicized.

Kelly	Harassment is a little extreme - you're the one who always posts on my talk page, I don't recall if I've ever posted on yours.	Jossi	HCon-
Kelly	I learned about the RfA from a friend on Commons, where I am a frequent contributor.	Jossi	
Jossi	OK. I will. Just play nice, that is all I ask.	Kelly	

Kelly's utterance in line 2 annotated as SCon- by virtue of it being a confession.

Criteria for annotation

Line 1

HAut-: orders, advice

HCon-: expressions of disapproval, criticism, or ridicule; expressions which are patronizing or condescending towards H

Line 2

SCon-: confession

Line 3

HCon-: expressing criticism, disapproval

[SAut-] Speaker autonomy face damaging

Definition

A discourse action which causes H (not necessarily the addressee) to think of S as more obligated to act in a particular way (different from what S would likely wish).

Criteria

1. Those actions that cause S to become more obligated to act in a particular way (different from what the S would likely wish).
    (a) expressing thanks (S accepts a debt)
    (b) acceptance of H's thanks or H's apology (S may feel constrained to minimize H's debt or transgression, as in "It was nothing, don't mention it.")
    (c) acceptance of offers (S is constrained to accept a debt, and to encroach upon H's autonomy face)
    (d) undesirable promises, commitments (to do or not to do something), and offers (*S commits themselves to some future action*, as witnessed by H)

Example

WikiTalk

Speaker	Utterance	Hearer	Face Effect(s)
Kelly	What's that supposed to mean?	Jossi	HCon-, HAut-
Jossi	You answer the question I asked you first in <LINK>, and then I will oblige with a response to yours.	Kelly	HAut-, **SAut-**
Kelly	How about we just drop the whole thing?	Jossi	

Jossi's proposal is annotated as SAut- because, while they were the one who suggested they would provide a response to Kelly, their use of *oblige* in "I will oblige with a response" signals some unwillingness to do so.

Criteria for annotation

Line 1

HCon-: challenges

HAut-: requests

Line 2

HAut-: requests

SAut-: undesirable promises, commitments

[HCon-] Hearer connection face damaging

Definition

A discourse action which causes H to believe that another (possibly S) thinks of H differently than H would likely wish.

Criteria

1. *Those that show that S has a negative evaluation of some aspect of H's connection face:*
   - *(a) expressions of disapproval, criticism, contempt or ridicule, complaints and reprimands, accusations, insults (S indicates that S doesn't like/want one or more of H's wants, acts, personal characteristics, goods, beliefs or values)*
   - *(b) contradictions or disagreements, challenges (S indicates that they think H is wrong or misguided or unreasonable about some issue, such wrongness being associated with disapproval)*

(c) expressions which are patronizing or condescending towards H (S indicates they think H is inferior in some way)

(d) poorly received actions associated with HCon+, apologies (compliments, agreement, understanding, etc. which are met with offense – S offends H by demonstrating in some way that they view H differently than they would like) – "You're pretty good at programming for a girl. - What's that supposed to mean?"

- This may be accompanied by SCon-.

2. *Those that show that S doesn't care about (or is indifferent to) H's connection face:*

(a) expressions of violent (out-of-control) emotions (S gives H possible reason to fear S or be embarrassed by S)

(b) *irreverence, mention of taboo topics, including those that are inappropriate in the context (S indicates that S doesn't value H's values and doesn't fear H's fears)* – this may be accompanied by HAut- should this impose discussion of the topic)

(c) bringing of bad news about H, or good news (boasting) about S (S indicates that S is willing to cause distress to H, and/or doesn't care about H's feelings)

(d) raising of dangerously emotional or divisive topics, e.g. politics, race, religion, women's liberation if met with resistance, reluctance, trepidation, etc. (S raises the possibility or likelihood of face-threatening acts (such as the above) occurring; i.e., S creates a dangerous-to-face atmosphere)

(e) *blatant non-cooperation in an activity – e.g. disruptively interrupting H's talk, making non-sequiturs or showing non attention (S indicates that they don't care about H's positive-face wants)* – this is likely also accompanied by HAut- as it prevents H from speaking)

(f) mis*use of address terms and other status-marked identifications (S may misidentify H in an offensive or embarrassing way, intentionally or accidentally)*

Example

TOWIE_1

Speaker	**Utterance**	**Hearer**	**Face Effect(s)**
Chloe	you're trying to paint yourself up to be this .h lovely ⌜little⌝ ((ges- tures as if picking up something delicate))	Meg	**HCon-**, HAut-
Meg	⌊no I'm⌋ no-=	Chloe	**HCon-**, HAut-
Chloe	=yeah you a:re	Meg	

Criteria for annotation

Line 1

HCon-: expressions of disapproval, criticism, contempt or ridicule, complaints and reprimands, accusations, insults

HAut-: expressions of strong (negative) emotions toward H

Line 2

HCon-: contradictions or disagreements, challenges

HAut-: interruption

[HAut-] Hearer autonomy face damaging

Definition

A discourse action which causes H to believe that H is more obligated to act in a particular way (different from what H would likely wish).

Criteria

1. *Those actions that predicate some future action A of H, and in so doing put some pressure on H to do (or refrain from doing) action A.*
    (a) *orders and requests (S indicates that they want H to do, or refrain from doing, some action A)* including those which are indirect (e.g., fishing for compliments, name-dropping, etc.)
    (b) suggestions, advice (S indicates that they think H ought to (perhaps) do some action A)
    (c) remindings (S indicates that H should remember to do some action A)
    (d) threats, warnings, dares (S indicates that they - or someone, or something - will instigate sanctions against H unless he does A)
    (e) complaints (S indicates that H should provide a justification or explanation for the issue that S raised)
2. *Those actions that predicate some positive future action of S toward H, and in so doing put some pressure on H to accept or reject them, and possibly to incur a debt:*
    (a) offers (S indicates that S wants H to commit themselves to whether or not H wants S to do some action for H, with H thereby incurring a possible debt)
    (b) promises (S commits themselves to a future action for H's benefit)
3. *Those actions that predicate some desire of S toward H or H's goods, giving H reason to think that H may have to take action to protect the object of S's desire, or give it to S:*
    (a) compliments, expressions of envy or admiration (S indicates that they like or would like something of H's)
    (b) expressions of strong (negative) emotions toward H – e.g. hatred, anger, lust, coveting (S indicates possible motivation for harming H or H's goods)

4. Those actions that by their nature physically prevent H from performing an action A typically available to H.
    (a) interruptions, shouting, laughing, speaking (H cannot converse if S is speaking at the same time)
    (b) leaving the conversation, other antisocial behavior (H cannot carry out a conversation which no longer has an audience)

Example

TOWIE_1

Speaker	Utterance	Hearer	Face Effect(s)
Meg	⌊wha-⌋ ve I been up to in Marbella	Chloe	**HAut-**, HCon-
Chloe	£ah don't know£ you tell me=	Meg	**HAut-**, SCon-, HCon-
Meg	=you tell me you've been tellin' everyone,= ((smiles))	Chloe	**HAut-**, HCon-
Chloe	=s just wha- i saw=	Meg	

Criteria for annotation

Line 1

HAut-: request (for answer); interruption

HCon-: expressions which are patronizing or condescending towards H

Line 2

HAut-: order/request

SCon-: acting stupid, apparent insincerity in actions

HCon-: blatant non-cooperation in an activity

Line 3

HAut-: order/request

HCon-: accusation

Annotation protocol

1. **Read the source utterance and its reaction.**

2. **For each of the four face damaging effects, identify the presence of each individually.** Refer to the guidelines in the annotation manual in order to determine what constitutes presence. While the definition of each face effect is more broad, it is ultimately what should be referred to in deciding whether to annotate the face effect or not. The Criteria section provides more detailed guidelines, including examples of face effects that are conventionally associated with certain actions. Therefore the guidelines in this section will not always be representative of the data to be annotated.

3. **Revise annotations when context reveals that the effect differs from what was first annotated.** Nevertheless, keep in mind what the interlocutors know at any given time in the dialogue.

Bibliography

Amido, S. (2025). *You're offended, I'm offended: A face-based analysis of confrontational conversation*. PhD thesis, Universitat Pompeu Fabra.

Amido, S. and Soubki, A. (2024). Towards an annotation of face effects in dialogue. Oral presentation at Grup de Linguïstica Formal (GLiF) Seminar, Universitat Pompeu Fabra, Barcelona.

Amido, S. and Soubki, A. (2025). Annotation of face effects as a tool for valence analysis. Oral presentation at EPITHETS & STAL Workshop 2025, University of Genoa, Genoa.

Brown, P. and Levinson, S. C. (1987). *Politeness: Some Universals in Language Usage*. Cambridge University Press.