The link between syntax, semantics, discourse, and lexicon in counteridenticals

A multivariate extension of co-varying collexeme analysis

Jesús Olguín Martínez[1] and Stefan Th. Gries[2,3]

[1] University of Hong Kong | [2] UC Santa Barbara | [3] JLU Giessen

The present study goes beyond traditional usage-based work in that it pays close attention not only to the interaction of lexicon and syntax in language use, but also to how other analytic layers of analysis (e.g., discourse) can influence the compatibility of lexemes in particular slots of constructional schemas. To investigate this domain, we examine counteridentical constructions (e.g., *if I were you, I would do it*) in a dataset of more than 1,000 examples from The Corpus of Contemporary American English. We focus on significant interdependencies between the slots of the protasis (i.e., types of NPs appearing in the protasis) and apodosis (i.e., semantics of the verb lemma) and how these cross-clausal associations interact with other linguistic variables such as the time reference of the apodosis, the discourse function of the construction, and the order of the protasis and apodosis. We demonstrate a novel application of a multivariate extension of co-varying collexeme analysis via a hierarchical configural frequency analysis.

Keywords: counteridentical, counterfactual, conditional, collexeme analysis

1. Introduction

Humans often anticipate what another will do by, in imagination, "placing themselves in the other's situation and simulating the other's mental economy" (Pelletier 2004:139–140). Such mental acts are often linguistically encoded in COUNTERIDENTICAL CONSTRUCTIONS, as in (1).

(1) [protasis If I were you], [apodosis I would have arrived on time].

This is a construction in which the *if* clause, the PROTASIS, identifies two inherently incompatible entities with each other (Goodman 1991). In (1), even though

the speaker and the addressee are not the same person, there exist certain worlds in which they are related via an equivalence relation (Kauf 2017: 43). Put another way, the speaker transposes herself into the world of the addressee (Arregui 2019; Lakoff 1996).

In these constructions, the protasis is counterfactual for the pragmatic reason that, in (1), the speaker is not the addressee in the actual world (Declerck & Reed 2001: 100), which is illustrated in (2) and (3) (from The Corpus of Contemporary American English (COCA)).

(2)   a.   If I were you, and, thankfully, I am not, I'd spend the next year taking a remedial English class (Bk:OnlyUni).
      b.   oh, my friend, if I were you, and I'm glad I'm not, I'd prepare myself for the inevitable (Oscar).
      c.   So if I were you, which I'm not, I would be bored, too (Austin Found).

(3)   **Speaker A:** Wait, wait. I wouldn't do that if I were you.
      Speaker B: Well, you're not me, are ya? (Baywatch).

Studies on counteridenticals have focused only on individual languages and on a handful of relatively isolated features of this complex sentence construction, which we discuss now.

First, a number of studies have explored the discourse functions of this construction and two main functions can be distinguished. Both formal and usage-based literature have shown that some counteridentical constructions serve the function of giving advice to the interlocutor (Kauf 2017; Van linden 2021: 278). For instance, (4) means 'if I imagine being you (identity of speaker and addressee), what I would do in your place is run every day'.

(4)   If I were you, I would have run every day (Invasion).

However, there is also another function that counteridentical constructions can serve, namely an imaginative function (Lewis 1973). This construction must not be understood as advice but as 'if I imagine being like you (similarity between speaker and addressee in terms of physical qualities, emotional states)…' Accordingly, (5) means 'if I were like you, I would be happy, too'. The example in (6), too, has an imaginative function and means 'if I were fast like Usan Bolt, I would be able to run 100 meters in just 9.58 seconds, too'.

(5)   If I were you, I would be happy, too (Revenge).

(6)   If I were Usan Bolt, I would be able to run 100 meters in just 9.58 seconds (Mov:Rock).

Second, other studies have focused on the lexical preferences of the main clause of this construction, the APODOSIS, which are relatable to the two functions just discussed. One of the key characteristics of the first type of counteridenticals, the advice-giving constructions, is that properties which can be advised need to be properties a person can influence (Kauf 2017: 16). Accordingly, activity predicates are expected to occur in this type of construction, as in (7a). On the other hand, the second type of counteridenticals, the imaginative ones are expected to appear with verbs denoting ability (7b) (Kauf 2017: 16).

(7) a. If I were you, I wouldn't touch it (Sliders).
    b. If I were you, I would be able to hike (Short Time).

Third, it has been argued that there is a relation between the time reference of the apodosis and the counteridentical's function. Specifically, an advisability reading of counteridenticals has been argued to only arise when the apodosis is interpreted as referring to the future, as in (8a), since it seems pragmatically odd to advise someone to do something in retrospect, as in (4) (Declerck & Reed 2001: 272). The illocutionary function of a piece of advice is that of getting the addressee to do something in the (immediate or distant) future. Accordingly, advice counteridentical constructions found in discourse contexts involving a past situation, as in (8b), are expected to be rare (Declerck & Reed 2001: 273). On the other hand, imaginative counteridenticals can occur not only in contexts involving future, but also in contexts involving present and past temporality.

(8) a. If I were you, I wouldn't do that (WTF!).
    b. If I were you, I would have studied harder. You need to focus on your goals (I See You).

Fourth, the types of pronouns appearing in the protasis of counteridenticals have also received attention (e.g., Thomas 2008). It is commonly assumed that advice can only be given in a speaker-addressee-context and, thus, should have the form: *if I were you,...* On the other hand, such a constraint does not seem to hold in imaginative counteridenticals in that they can be found in both a speaker-addressee-context or another type of context: *if I were X,...* where X is not the addressee of the conversation (e.g., 3rd person pronouns or noun phrases referring to a 3rd person).

Finally, besides these key formal properties of counteridentical constructions identified in the literature, there seems to be another feature that is also relevant to the analysis of this construction, i.e., the relation of the counteridenticals' function to the order of the protasis and apodosis. Protases and apodoses tend to follow an iconic order (e.g., *if you study hard, you will pass the exam*). Put another way, protases tend to precede apodoses, because the protasis refers to a situation

that is conceptually, or even logically, prior to the one expressed in the apodosis (Diessel 2008: 469). However, we also know that apodoses may not follow this iconic order (e.g., *you should go if you ask me*). In this scenario, they have been characterized as clauses with a pragmatic function, i.e., clauses that refer to linguistic or non-linguistic aspects of the ongoing social interaction/speech situation itself (Hampe & Gries 2018).[1] The function of advice-giving counteridenticals is based on the speaker's confidence and personal (emotional) experience about the benefits of the advice for the advisee. Given that these constructions are based on the speaker's personal (emotional) stance on the proposition, they all concern dimensions in which the status of a situation is at stake, e.g., in terms of whether it is agreeable or not. This may provide an explanation as to why these constructions typically attract speaker-interlocutor interaction (see Nuyts 2015: 110 for similar claims with respect to other constructions, such as complement constructions denoting mental states). Accordingly, apodoses in advice counteridentical constructions should precede their protases whereas apodoses in imaginative counteridenticals follow their protases.

While such studies have provided some description of counteridentical constructions, we know next to nothing about how any of the features sketched above interact with one another in counteridentical constructions; this is precisely the aim of the present study. We focus on significant interdependencies between the slots of the protasis (i.e., types of NPs appearing in the protasis) and apodosis (i.e., semantics of the verb lemma) and how these cross-clausal associations interact with other linguistic variables such as the time reference of the apodosis, the discourse function of the construction, and the order of the protasis and apodosis. In very general terms, the question is, 'can we identify preferred and dispreferred co-occurrences of the features mentioned before?', but the more specific question is whether the patterns of preferred co-occurrences of features that one would infer from the above-discussed previous work and that are summarized in Table 1 are indeed attested especially if one looks at this in a multivariate way.

Our exploration of these expectations is grounded in a usage-based Construction Grammar approach. That is, we assume there are associative connections between individual lexemes and specific slots of constructions: the distributional biases and constraints of lexemes and constructions are not arbitrary, but functionally motivated or even predictable. This is in essence Goldberg's (1995: 50) SEMANTIC COHERENCE PRINCIPLE, according to which constructions attract lexical items that are compatible with the semantic specifications of particular slots.

---

1. This seems to match speech-act conditionals in Sweetser's (1990) three-way distinction of this complex sentence construction. In her proposal, Sweetser mentions that conditionals can be characterized as content, epistemic, and speech-act conditionals.

Table 1. Summary of expectations

CXTYPE	NPTYPE1	NPTYPE2	APODOSISLEMMA	TIME REFERENCE	POSITION
advice	*1st person*	*2nd person*	*activity verbs*	*future*	*preposed apodosis*
imaginative	*1st person*	*2nd person, 3rd person*	*ability verbs*	*past, present, and future*	*postposed apodosis*

The present work goes beyond this traditional work and demonstrates how considering multiple linguistic features at the same time can provide a more holistic analysis of filler-slot relations.

The remainder of the paper is structured as follows. In Section 2, we discuss our corpus data and the variables for which our counteridentical constructions have been annotated, i.e., NPs of the protases, apodosis lemma, discourse function, time reference, and position of clauses; also, we introduce how we statistically analyze the interaction of syntax, semantics, discourse, and lexicon in this complex sentence construction. In Section 3, we present our results and, on the basis of these results, Section 4 provides some discussion regarding preferred and dispreferred co-occurrences of the features taken into account here. Moreover, we explain that such clusters can be explained by different semantic factors. The findings of Section 4 then lead to Section 5, where we show that they have important methodological and theoretical implications to our general understanding of filler-slot relations. Section 6 offers concluding remarks and points to issues that remain to be investigated by future studies.

2. Methods

2.1 Corpus data

We exhaustively retrieved a sample of counteridentical constructions by searching COCA for the form *if I were*,[2] and then manually inspecting the resulting candidate hits to identify false positives, thus, for instance, excluding constructions with the following patterns in (9)–(11):

(9)   The man acted as **if I were** a stranger.

---

2. Since these constructions are mostly attested in conversational settings, most matches came from TV and movies subtitle data.

(10)   Even **if I were** staying, I'm not going back to school.

(11)   **If I were** doing evil, I'd be scared stiff.

The resulting data sample of 1,007 true hits was then annotated for the following variables:

- CXTYPE: whether counteridentical constructions are used for expressing advice or have an imaginative function;
- NPTYPE: type of NP that occurs after the copula *were* in the protasis;
- APODOSISLEMMA: verbs that can occur in the slot of the apodosis;
- TIME REFERENCE: the temporal reference of the counteridentical construction;
- POSITION: whether the apodosis appears before (preposed) or after the protasis (postposed).

Table 2 exemplifies the structure of our data frame. In the following sections, we describe the variable levels and their annotation in more detail.

Table 2. Illustration of the data annotation in the present study

SOURCE	EXAMPLE	NPTYPE	APODOSISLEMMA	CXTYPE	TIME REFERENCE	POSITION
The Orville	*If I were you, I'd make a break for it*	2nd pers.	*make* (activity)	advice	future	postposed
Banshee	*If I were a bird, I would have been able to fly*	3rd pers.	*be able to* (ability)	imaginative	past	postposed
NorthAmRev	*I wouldn't talk to her if I were you*	2nd pers.	*talk* (utterance)	advice	future	preposed
RedCedarRev	*I would walk slower now if I were you*	2nd pers.	*walk* (motion)	advice	present	preposed
mediaite.com	*If I were Mario, I would be nervous, too*	3rd pers.	*be* (state)	imaginative	present	postposed

2.1.1 *NPs of the protases*

Protases in counteridentical constructions can appear with different types of NPs (Thomas 2008). The present study only takes into account counteridentical constructions that are speaker-oriented (i.e., *if I were…*). Accordingly, we focused on the second type of NP, i.e., the NP appearing after the copula *were*. These NPs could be 2nd person pronouns, 3rd person pronouns, or full NPs referring to a 3rd person (e.g., *Michael, man, bird*). This means that counteridentical constructions occur in speaker-addressee-contexts, as in (12), and speaker-non-addressee-contexts, as in (13).

(12)   I would be more careful with what I say if I were you (B.T.K. Killer).

(13)   If I were a bird, I would be able to fly (Mov:Scream)

Note that instances in which the first-person pronoun does not appear in counteridentical constructions, as in (14)–(15), are not taken into account in the present study:

(14)   I wouldn't do that in your position.

(15)   I wouldn't drink that in your place.

2.1.2 *Apodosis lemmas*

We annotated each construction with regard to the lemma occurring in the apodosis of counteridentical constructions, but there also were examples of apodoses that did not occur with any verb lemma, as in (16).

(16)   I wouldn't if I were you (Crazy Ex-Girlfriend)

In this scenario, it was important to pay close attention to preceding stretches of discourse to identify the missing verb lemma of the apodosis. For instance, the construction in (16) was preceded in the context by the following construction: *You don't have to believe me.* Accordingly, in this scenario, we determined the verb lemma of the apodosis to be *believe.*

2.1.3 *Discourse function*

It has been shown that conditional clauses may serve different discourse functions, such as exploring options, providing polite requests, illustrating, contrasting, and assuming (Ford & Thompson 1986). Due to their discourse functions and interpersonal function in language, conditionals have been characterized as pragmatic conditionals (Sweetser 1990).

As mentioned in Section 1, counteridenticals are conditional constructions that can be used for advisory or imaginative discourse functions. Determining the

functions of counteridentical constructions was quite labor-intensive because it is not sufficient to analyze biclausal constructions as in (17) without taking into account their discourse context. For instance, without the discourse context of (17), it is not clear whether this construction should be characterized as having an advice-giving or imaginative function. However, by looking at preceding and/or subsequent stretches of discourse, it was possible to determine that the example in (17) has an advice-giving function, as can be seen in (18). Note that counteridentical constructions typically attract speaker-interlocutor interaction (Nuyts 2015: 110). It is this methodological step that enabled us to determine whether a construction had an advice-giving or imaginative function.

(17)   I would be there if I were you.

(18)   a.   Lionel told me to visit my family's physician as soon as possible.
       b.   I would be there if I were you.
       c.   I agree. I think it'd be a good idea (Smallville).

As for counteridenticals involving advice, we rely on two main discourse features. First, advice-giving actions expressed with counteridenticals may be solicited, as in (19). Accordingly, this factor played an important role in classifying this construction as advice-giving. Second, we also paid close attention to whether the addressee agreed or not with the advice offered, as in (20).

(19)   a.   Wait — shouldn't we do something about this?
       b.   If I were you, I'd get out of here as soon as possible (Mov:SleeplessSeattle)

(20)   a.   I wouldn't be talking about what was on my shirt, baby, if I were you.
       b.   You're right! (Fox: The Five)

As for counteridenticals indicating how similar a speaker would be to X in terms of physical qualities or emotional states, they never occurred in the two contexts mentioned above (i.e., asking for advice and/or agreement and disagreement). Rather, they co-occurred with utterances in which a speaker indicates how much he wishes to have similar physical characteristics to X, as in (21). In other instances, they were preceded by utterances in which a speaker discusses the feelings of someone else and then the addressee employs a counteridentical to express that he would feel in the same way in X's place, as in (22). This discourse context can be characterized as involving empathy in that one is able to stand in the shoes of another person, i.e., to understand the emotional situation they are in. Put another way, in this scenario, there is an affective response that stems from the apprehension or comprehension of an other's emotional state or condition (Preston & de Waal 2002).

(21)   Many women do that and they are absolutely the best. If I were a woman, I
        would be able to do that (lastpsychiatrist.com).

(22)   a.   Nate is mad at him.
        b.   I'm sorry that he has a stubborn unfriendly neighbor. If I were him, I
              would be upset, too (dk-wdyt.blogspot.com)

2.1.4   *Time reference*

It has been shown that counterfactual constructions can have past temporal ref-
erence when they express a conditional relationship between two situations that
failed to be realized (e.g., *if John had come yesterday, we would have had fun*;
Declerck & Reed 2001: 177). They can also have present and future time reference,
as can be seen (23)–(25). We have annotated each counteridentical construction
in our dataset in terms of their temporal reference.

(23)   If I were you, I would have done that (past temporal reference).

(24)   If I were you, I would be happy now (present temporal reference).

(25)   I wouldn't go there next week if I were you (future temporal reference).

2.1.5   *Position of clauses*

The variable positioning of English adverbial clauses with respect to their main
clauses is a domain that has received considerable attention in previous research
(e.g., Diessel 2008: 469). As was mentioned in Section 1, counteridentical con-
structions, as a subtype of conditional constructions, may appear in iconic (apo-
doses occur postposed to the protasis) or non-iconic orders (apodoses occur
preposed to the protasis), so we've annotated for this binary distinction. There
were only a couple of examples in which it was not clear whether the complex sen-
tence occurred in iconic or non-iconic order given that the protases were embed-
ded in the apodoses in these examples, as in (26)–(27). We decided to disregard
these examples from the present study.

(26)   I wouldn't, if I were you, touch that! (Law & Order).

(27)   I known it's none of my business, but I wouldn't, if I were you, go there! (Alien
        Nation).

In the following section, we outline the statistical approach we used to analyze the
data in more detail.

2.2   Statistical analysis

The perhaps most widely-used way in which corpus-based approaches have studied associative connections between individual lexemes and specific slots of constructions is by methods from the family of COLLOSTRUCTIONAL ANALYSES (CA). The family of methods of CA distinguishes three different approaches:

i.   COLLEXEME ANALYSIS, which quantifies how much words that occur in a syntactically defined slot of a construction are attracted to or repelled by that construction (Stefanowitsch & Gries 2003);

ii. DISTINCTIVE COLLEXEME ANALYSIS, which quantifies how much words prefer to occur in slots of two functionally similar constructions (Gries & Stefanowitsch 2004a);

iii. CO-VARYING COLLEXEME ANALYSIS, which quantifies how much words in one slot of a construction are attracted to or repelled by words in a second slot of the same construction (Gries & Stefanowitsch 2004b; Stefanowitsch & Gries 2005).

In the present study, we demonstrate a novel application of a multivariate extension of co-varying collexeme analysis via a HIERARCHICAL CONFIGURAL FREQUENCY ANALYSIS (HCFA). For this, we include all the annotated features described above at the same time.

HCFA is essentially a multivariate extension of chi-squared tests (von Eye 2002; Gries 2009: 240). The analysis generates all possible 1 to n variable frequency tables; in our case, we have 5 variables, which means that the analysis generates $n = 31$ tables: 5 1-dimensional tables, 10 2-dimensional tables (i.e., involving two variables), 10 3-dimensional variables, 5 4-dimensional variables, and 1 table with all 5 variables. For each cell of each table, a HCFA tests whether the specific combination of values — the so-called CONFIGURATION OF THE CELL — occurs more often or less often than expected (based on the overall frequencies of those values). The comparisons of each cell's/configuration's observed and expected value can involve significance tests and, in this case, we use exact binomial tests adjusted for multiple comparisons; because of the exploratory nature of this application, we explored the configurations with p-values of less than 0.1. In addition, we compute an effect size that expresses the size of the difference between observed and expected frequencies with a measure related to correlation coefficients, the Q_r coefficient. In the discussion below, we will follow the standard terminology of using the terms TYPES and ANTITYPES, for situations where obs>exp (i.e., a kind of attraction relation) and obs<exp (i.e., a kind of repulsion relation) respectively.

There are two motivations for using HCFA to analyze types and antitypes of counteridentical constructions. First, HCFA is an appropriate alternative to popular regression models due to the nature of our data. In a regression model, one or more independent variables are used to predict a dependent variable. However, our data do not come with an obvious division of the five variables into one response/effect and four predictors/causes, which is why a multivariate method that does not require the response-predictor dichotomy is better suited to the data. Second, a HCFA is both an exploratory and a hypothesis-testing method. It is exploratory "because every possible combination of variable levels gets tested for the presence or absence of an effect, and it tests hypotheses because each combination is subjected to a significance test" (Gries 2009: 241). Given that the analysis of types and antitypes of counteridentical constructions is exploratory, HCFA is ideal to investigate this domain; note, however, that (H)CFA can also be used confirmatorily, see Krauth (1993) and von Eye (2002) for discussion.

3.   Results

Since HCFAs generate a large number of tables of combinations of variable levels, they come with a multitude of results; in fact, here, the number of theoretically possible combinations of variable levels is 13,067 (with observed token frequencies in the interval [0, 916]). To facilitate the discussion of the results, we analyze types and antitypes by position of clauses; we will first discuss findings for the postposed construction and then for the preposed construction and for each construction we will discuss types and antitypes in prose and summarize the overall findings heuristically and diagrammatically.

3.1   Types and antitypes when position is postposed

Counteridentical constructions with postposed apodoses (e.g., *if I were you, I would do it*) have only a small number of very concentrated (in the sense of 'interrelated, mutually supportive') types. The by far strongest type configuration is that of {3rd person, imaginative, present time reference}: each of these features is a strong and significant type in isolation, but all their pairwise combinations are, too, as is the triplet of features (with a frequency of 79, a Q_r score of 0.276, and a type/attraction p_{adjusted}-value of $<10^{-90}$).

In addition, there are several strong and significant lexical preferences similar to, but much less comprehensive than, what traditional collostructional analyses might detect (see below for more discussion of such differences): postposed apodoses prefer especially *be*, *be able to*, and *want*, but the attractions of some of these verbs to postposed apodoses are even stronger when they are also used together with one or more features of {3rd person, imaginative, present time reference}. For example, *be able to* is attracted to postposed apodoses with a Q_r-value of 0.129, but if the usage also involves 3rd person or imaginative functions or present time reference, the Q_r-values increase to 0.158 and 0.153 or 0.15 respectively, indicating a more specific type than any mere monofactorial analysis could discover. Similarly, *be* is attracted to postposed apodoses with a Q_r-value of 0.368, but if *be* is coupled with 3rd person or imaginative functions, then the effect size changes to a lower 0.185 but also a higher 0.431 respectively; in fact, imaginative function and present time reference regularly increase the attraction of all these verbs to postposed apodoses, and both imaginative function and present time reference can make 2nd person also exhibit a notable preference for postposed apodoses.

Turning to antitypes, postposed apodoses do also have a variety of dispreferences, which are again mutually supportive and feature the characteristics of 2nd person, advice function, and future time reference, which also combine into a strong antitype with all three features combined. However, a closer analysis also indicates that both function and person actually do not matter: as soon as there is future time reference, both functions and both 2nd and 3rd person are significantly dispreferred in postposed apodoses (compared to, as in all cases, the expected frequencies for these combinations resulting from the marginal totals as is the case in, say, any chi-squared test). In terms of lexical (dis)preferences, *watch* is dispreferred, and *be* as well as *do* are dispreferred in a variety of closely-knit antitypes involving one or more features of {2nd person, advice or imaginative, future}.

Since types and antitypes can often be described in different linguistic ways, or from different vantage points, we follow Olguín Martínez & Gries (2024) and summarize the results in a plot reminiscent of network representations that have a long history in cognitive and usage-based linguistics. Figure 1 represents the results for postposted apodoses; Figure 2 shows the analogous results for preposed apodoses and, in both figures, types (i.e., combinations of levels attracted to post- and preposed respectively) and antitypes (i.e., combinations of levels repelled by post- and preposed respectively) are listed above and below the central nodes.
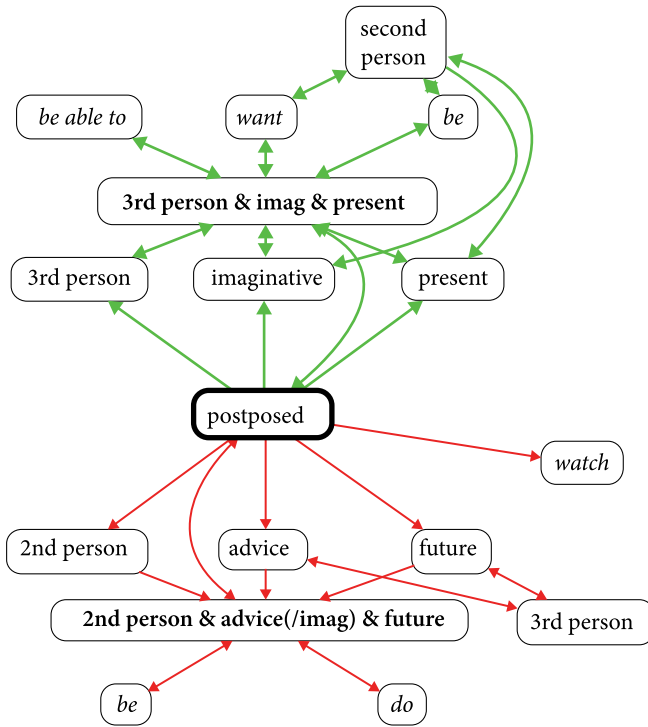
Figure 1. Types (green) and antitypes (red) for constructions with postposed apodoses

3.2 Types and antitypes when position is preposed

The results for preposed apodoses are numerous and reveal several complex type/antitype combinations that not only involve many lexical (dis)preferences but also express what generically would look like 'X likes Y but not [or only] when Z_1 or Z_2' which, we submit, are again often patterns that would be impossible to discern without some kind of multivariate analysis.

There are again strong preferences for several features as well as their joint configuration: preposed apodoses like to occur in/with 2nd person, future time reference, and advice function. In addition, there are many collexemes that prefer to occur with preposed apodoses and usually one or more of those structural features; *do*, *go*, *get*, and *watch* are particularly associated with preposed position and those features, while *touch*, *take*, and *worry* are more specialized and only show such a preference with future time reference and advice function.

The range of antitypes of preposed apodoses is much broader. Generally, these disprefer 3rd person, imaginative function, and present time reference and their combinations. In terms of verb lemmas, the verb lemma *be* exhibits such a

wide range of dispreferences that it is basically dispreferred with just about everything else. The verb lemma *go* is dispreferred especially with present time reference and in imaginative functions and *be able to* is dispreferred in general as well, but also particularly when used with 2nd person, future time reference, and in advice contexts.
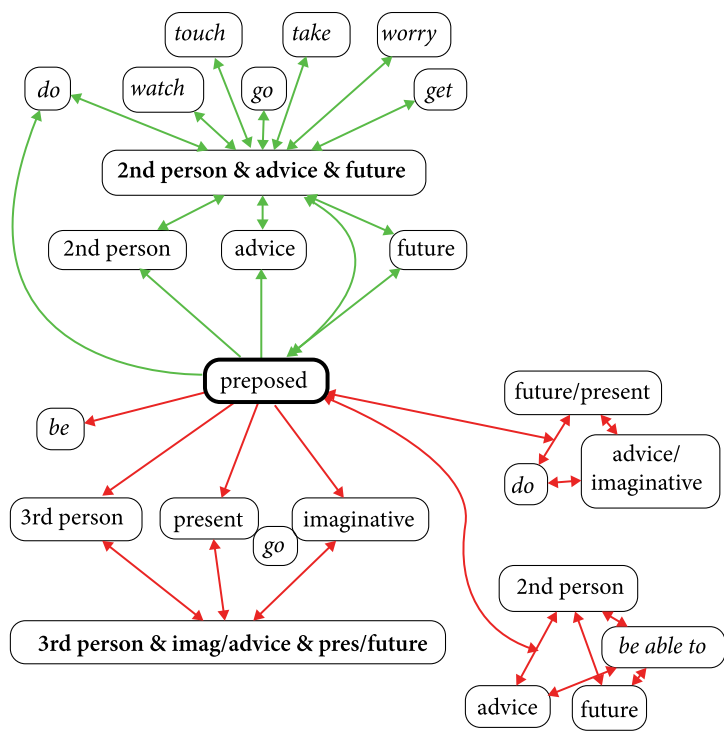

Figure 2. Types (green) and antitypes (red) for constructions with preposed apodoses

4.   Discussion

The results presented in Section 3 allow for several observations regarding counteridentical constructions in American English. In this section, we discuss these results and consider their linguistic implications.

4.1   Counteridentical constructions containing postposed apodoses

4.1.1   *Preferred co-occurrences*

As outlined in Table 1 in Section 1, we expected that counteridentical constructions containing postposed apodoses would (prefer to) occur with protases with 2nd and 3rd person, with apodoses with ability verbs (e.g., *be able to*), with past, present or future temporality, and in imaginative discourse contexts. This type is partially captured in the results of the present study in that it consists of {postposed apodoses, 3rd person, *be able to*, present temporal reference, and imaginative discourse contexts}, as in (28).

(28)   If I were him, I would be able to walk like he's doing it now (Bk:SecondBiteAt-Apple).

The fact that the verb lemma *be able to* co-occurs with postposed apodoses and in imaginative discourse contexts can easily be explained from a usage-based perspective. The verb lemma *be able to* can be characterized as a verb denoting that someone has the skills or qualities (i.e., ability) to carry out an action. This verb lemma is not suitable for discourse contexts that have a framing potential that is relevant to the organization of ongoing social interaction (preposed apodoses), i.e., involving a recommendation, suggestion, or advice. Put another way, given that postposed apodoses do not display the framing potential that preposed apodoses have (see Section 4.2), this may provide an explanation as to why *be able to* likes to occur in postposed apodoses in counteridentical constructions, as in (28). Note that in this type, there is a similarity between the speaker and X in terms of physical qualities. In this scenario, X is not the addressee of the conversation. This means that the speaker is not telling X what to do if he were in a similar situation (Couper-Kuhlen & Thompson 2022). Accordingly, 3rd person is a natural fit to this cluster.

As for the other two types with postposed apodoses identified in Section 3, they also align with our expectations. The first type, {postposed apodoses, 3rd person, *be*, present time reference, and imaginative function} as illustrated by (29), can similarly be explained from a usage-based perspective. This type also brings further support to our initial expectations in that imaginative counteridentical constructions with postposed apodoses are preferred given that they do not have a framing potential that is relevant to the organization of ongoing social interaction. Rather, they are only used to express similarity between speaker and third person entities in terms of physical qualities and emotional states. The verb lemma *be* harmonizes with the meaning of this construction in that it indicates traits or features of a person's body or person's feelings and moods.

(29)   If I were Jack, I would be happy, too. He must be proud of his son
       (CNN_King).

The second type, {postposed apodoses, 3rd person, *want*, present time reference, and imaginative function} as in (30), is also in line with our expectations. Counteridenticals with an imaginative function are a natural fit to postposed apodoses given that they do not involve a key organizational role of an ongoing social interaction, e.g., a recommendation, suggestion, or advice, as preposed apodoses do. One comment on the apodosis verb lemma *want* is in order here. Although imaginative counteridenticals in this type do not appear with verbs denoting ability, as expected in the present study and previously shown by other studies (e.g., Kauf 2017:16), the verb *want* aligns with the semantics of this construction given that it is a verb that profiles the desire for a situation to obtain (i.e., emotional state). Recall that in imaginative constructions there exists a similarity between speaker and addressee in terms of emotional states.

(30)   If I were him, I wouldn't want to be in his situation now. I wouldn't be happy
       about it! (ALF).

4.1.2   *Dispreferred co-occurrences*

The antitypes of counteridentical constructions containing postposed apodoses are fairly diverse, but suggest a number of interpretively useful patterns. First, counteridentical constructions containing postposed apodoses dislike 3rd person, future time reference, and advice function, as in (31). From a usage-based perspective, this dispreferred co-occurrence of features is sound: advice counteridentical constructions can only be given in a speaker-addressee-context (2nd person) and occur with preposed apodoses due to the framing potential that is relevant to the organization of ongoing social interaction (see Section 4.2). Note that the fact that this antitype disprefers future temporality has to do with the fact that the imaginative function may arise in different temporal spheres. However, an advisability reading of counteridentical constructions can only arise when the apodosis is interpreted as referring to the future (Declerck & Reed 2001:272). This may provide an explanation as to why future temporal reference is dispreferred in counteridentical constructions containing postposed apodoses.

(31)   If I were Mark, I would go next week to the annual meeting (Legit).

Second, another antitype deserving attention is that in which postposed apodoses dislike the cluster {2nd person, *be*, future time reference, imaginative function}, as in (32).

(32)   If I were you, I would be happy, too! (Madhouse).

Intriguingly, this co-occurrence of features, at first glance, aligns with the semantics of imaginative counteridentical constructions. This stems from the fact that imaginative counteridentical constructions express similarity between speaker and addressee in terms of physical qualities and emotional states (e.g., *be fast*, *be happy*, *be upset*). However, a closer look at the data reveals that it is not this function of *be* that is repelled from this construction. Rather, what is dispreferred from this cluster are instances like (33). In this scenario, the verb lemma *be* does not express a physical quality. Instead, it indicates a suggestion about what someone should do.

(33)   If I were you, I'd be studying for the test. It's likely you'll fail if you don't study (tor.com).

Third, postposed apodoses do not like the cluster {2nd person, *do*, future time reference, imaginative function}, as in (34). This dispreferred co-occurrence of features can be easily explained. The verb lemma *do* is a generic verb that involves activity: 'if I imagine being you (identity of speaker and addressee), what I would do in your place is…' Given its non-specific semantics, speakers prefer verb lemmas with more specific semantics in imaginative contexts.

(34)   If I were you, I would do exactly what you plan to do next week (justthink .org).

Fourth, the last antitype with postposed apodoses identified here is that in which postposed apodoses do not like to occur with the verb lemma *watch*, as in (35). The main reason why *watch* is repelled from this construction has to do with the fact that *watch* basically serves to indicate some situation which the speaker considers to be harmful and which should be avoided. As will be shown in Section 4.2., the semantics of this verb lemma aligns with counteridentical constructions used for giving advice.

(35)   If I were you, I would watch what I say in front of Aaron (renibel.net).

4.2   Counteridentical constructions containing preposed apodoses

4.2.1   *Preferred co-occurrences*

The strongest attraction of preposed apodoses is, as mentioned in Section 3.2, to the collexemes *touch*, *watch*, *get*, *take*, *worry*, and *do* when protases occur with 2nd person, with future temporal reference, and in advice contexts. This fits our initial expectation in which we predicted that preposed apodoses will (prefer to) appear with the cluster {2nd person, activity verbs, future time reference, and advice function}. Although the collexemes *touch*, *watch*, *get*, *take*, *worry*, and *do*

can be characterized as activity verbs, each of them deserves some comments in its own right.

First, the type {preposed apodosis, 2nd person, *touch*, future time reference, and advice} can be characterized as a construction used in the expression of a precautioning situation: the apodosis refers to some precautions that can be taken to avoid a potential harmful consequence involving physical damage. In particular, a speaker warns the addressee about a possible danger in case he brings a body part into contact with something, as in (36)–(38).

(36)   I wouldn't touch that if I were you. You might get burned! (Avengers of Justice: Farce Wars).

(37)   I wouldn't touch that object if I were you. You'll get hurt! (Hellraiser: Hellworld).

(38)   I wouldn't touch it if I were you. That's Yun's sword (Warriors of Virtue).

The verb lemma *touch* (appearing in protases with negative polarity in our corpus) denotes a state of alertness to exercise caution and, in this scenario, aligns with the semantics of advice counteridentical constructions with precautioning semantics.[3] Given its discourse relevance or deontic strength (in the sense of Couper-Kuhlen & Thompson 2022), this may provide an explanation as to why apodoses are preferred preposed to the protasis in this communicative scenario: *don't touch that to prevent X from occurring*. Precautioning clauses can be situated within the domain of apprehensives. Lichtenberk (1995:297) shows that the apprehensional domain must be understood as emotions triggered by an undesirable, (highly) possible situation. For the most part, studies dedicated to the analysis of this domain have paid close attention to negative purpose clauses (e.g., *hold of that child lest he fell*), *in case* clauses (e.g., *in case I die, I tell you first*), and timitive clauses (e.g., *I was afraid at night that they would harm me*; Schmidtke-Bode 2009:130; Smith-Dennis 2021).

Second, the type {preposed apodosis, 2nd person, *watch*, future time reference, and advice} can also be characterized as a construction used in the expression of a precautioning situation and in particular in contexts where, if the precaution expressed in the apodosis is heeded, the potential danger that may result in physical damage is avoided: *Y do not do X otherwise Z will happen*, as

---

3. As correctly pointed out by one anonymous reviewer, it looks like in some configurations the apodosis is preferably negated. While we acknowledge that this may be a morphosyntactic feature relevant to the analysis of types and antitypes of counteridenticals, we have decided to focus only on those features of counteridenticals for which there is some discussion on the literature.

in the examples in (39)–(41). In this scenario, *watch* must be understood as *be alert*, *be vigilant*, or *be careful*. It admonishes someone to keep a protective eye on something. Accordingly, it harmonizes with the semantics of advice counteridentical constructions that denote a precaution (as was discussed in Section 4.1.2, postposed apodoses do not like to occur with the verb lemma *watch*; see (35)).

(39)   I would watch what I say if I were you. He's very dangerous and may hit you (The Rocketeer).

(40)   I would watch my step if I were you. You may fall and get hurt (Omni).

(41)   I would watch my words if I were you. She may slap you (Revenge).

Third, the type {preposed apodosis, 2nd person, *get*, future time reference, and advice} is also a construction with a precautioning function. However, in this scenario, the speaker does not warn about a possible danger that may result in physical damage if the addressee does not carry out the precaution expressed in the apodosis. It just serves to indicate some situation which the speaker considers to be unpleasant, and which should be avoided, as in (42). The verb lemma *get* aligns with the semantics of advice counteridentical constructions that denote a precaution given that *get*, in this scenario, has a protective sense indicating 'obtain X to avoid an unpleasant situation'.

(42)   I would get another car if I were you. People will make fun of you if you drive that old car (Atlanta).

Fourth, the type {preposed apodosis, 2nd person, *take*, future time reference, and advice} is similar to the one discussed before in that it indicates that the apodosis situation should or should not be carried out in order to avoid a consequence that the speaker deems as undesired, as in (43). In fact, the verb lemmas *get* and *take* form a semantically related cluster of verbs in that they capture the precautious nature of the situation in the apodosis: 'obtain X to avoid an unpleasant situation'.

(43)   I would take an umbrella if I were you. I heard it will rain, so you don't want to get wet (Houston).

Fifth, unlike the types discussed so far, the type {preposed apodosis, 2nd person pronoun, *worry*, and future time reference} does not result in physical damage or an undesirable situation if the apodosis situation is not carried out. Rather, it indicates the speaker's attitude concerning the undesirability of the situation encoded in the apodosis (Lichtenberk 1995: 291). In (44), the pattern highlights the semantics of *worry* by portraying an entity X or situation as feared. Palmer (2001: 133–134) suggests that fear expressions indicate "little more than an unwel-

come possibility" or epistemic uncertainty. Following Givón (2001: 52–53), the construction in (44) involves epistemic anxiety in that it indicates the unresolved certainty of a situation.

(44)   I would worry about your final exam if I were you. I heard it'd be difficult (Northern Exposure).

Sixth, another type that deserves careful attention is the following {preposed apodosis, 2nd person, *do*, and future time reference}. This cluster is particularly interesting in that the apodosis may refer not only to some precautions that can be taken to avoid a potential harmful consequence involving physical damage, as in (45), but also to some situation which the speaker considers to be unpleasant, and which should be avoided (no physical damage involved), as in (46) (as was discussed in Section 4.1.2, postposed apodoses do not like to occur with the verb lemma *do*, see (34)).

(45)   Don't tell him that. I wouldn't do that if I were you. You might get hurt! (Dead Man).

(46)   Taking a nap? I wouldn't do that if I were you. You'd better stay and study as much as possible so that you don't fail the test (Senior Trip).

The verb lemma *do* can be characterized as a generic verb, whose meaning can only be recovered from a specific discourse context. Based on the discourse context of the example in (47), the verb lemma *do* refers to the following situation: *I wouldn't tell him that if I were you*. This indicates that the semantic characteristics of generic verbs can be seen to arise at the level of the construction or discourse rather than solely at the level of the generic verb's lexical semantics (Family 2011: 9). We argue that the generic verb *do* is used in this type to avoid the lexical repetition of the verb already provided in the previous context. Repetition avoidance in lexical/constructional choice may be driven by economy.

(47)   a.   Mr. Jones is complaining about the weather one more time.
       b.   I will tell him to stop doing that.
       c.   I wouldn't do that if I were you (Barb Wire).

4.2.2   *Dispreferred co-occurrences*

There are three antitypes of counteridentical constructions containing preposed apodoses that deserve closer attention. First, as was mentioned in Section 3, preposed apodoses do not like to appear with the cluster {2nd person, *be able to*, future time reference, and advice function}. As we argued in Section 4.1.1, the verb lemma *be able to* can be characterized as a verb denoting that someone has the skills or qualities to carry out an action. This verb lemma is not suitable for dis-

course contexts that have a framing potential that is relevant to the organization of ongoing social interaction (preposed apodoses), i.e., involving a recommendation, suggestion, or advice. This may provide an explanation as to why the verb lemma *be able* does not like to occur in preposed apodoses in counteridentical constructions with an advice-giving function.

Second, preposed apodoses do not like to appear with the cluster {3rd person, *be*, present time reference, and imaginative function}. As was discussed in Section 4.1.2, the verb lemma *be* is used in the expression of a similarity between speaker and addressee in terms of physical qualities and emotional states (e.g., *be fast*, *be happy*, *be upset*). In particular, this function is attested in imaginative counteridentical constructions in which the apodosis is postposed to the protasis. Given that preposed apodoses display a framing discourse potential, this may explain why the verb lemma *be* is repelled from counteridentical constructions containing preposed apodoses.

Third, the verb lemma *go* is dispreferred especially with the cluster {preposed apodoses, present time reference, and imaginative function}. The verb lemma *go* expresses the agent's moving from one place to another. Accordingly, it is not suitable for discourse contexts indicating similarity between the speaker and X in terms of physical qualities and emotional states. This may explain why *go* is repelled from imaginative counteridentical constructions.

5.   Methodological and theoretical implications

5.1   Methodological implications

One major advantage of the methodology employed here is how it goes beyond what has been the default kind of analysis for cases like the one studied here, i.e., a co-varying collexeme analysis. Our paper demonstrates as a methodological contribution that, in essence, regular collexeme analysis needs to be augmented with more distributional/corpus data/annotation than has been the standard so far. For example, a standard co-varying collexeme analysis would lose a high amount of information because it would only be able to consider two variables/columns within one and the same construction. More concretely and given that we structured our discussion above by position, this would mean that the default analysis might require of the analyst:

– one co-varying collexeme analysis of, say, CXTYPE and LEMMA when position is preposed;
– one co-varying collexeme analysis of, say, CXTYPE and LEMMA when position is postposed.

This would already be two of the old default kind of analyses. However, two analyses would then by definition not be able to see what, if any, types or antitypes might exist that involve TIME REFERENCE and/or PERSON. In other words, any of the results above that involve time reference and person could not be obtained that way.

In addition, there is another, more damaging shortcoming: in an attempt to approximate anything multi-dimensional, we just said that an analyst might have to do multiple standard co-varying collexeme analyses, e.g., a separate one for each level of position. But, if one does that, note what happens: the results for both analyses — for the preposed and the postposed positions — indicate that the following two types are among the top 3 types: {advice, *do*} and {advice, g*o*}, but that also means that such analyses by definition also cannot reveal which position {advice, *do*} or {advice, g*o*} prefers. Put differently, we know that, if one only looks at position: preposed, {advice, *do*} is a type, and if one only looks at position: postposed, {advice, *do*} is a type, but what an analyst cannot know from that is whether {advice, *do*} significantly prefers one of the two positions (more). Our truly multidimensional analysis, however, shows that {advice, *do*, preposed} is hugely significantly preferred ($p_{\text{adjusted binomial}} < 10^{-10}$) while {advice, *do*, postposed} is significantly dispreferred ($p_{\text{adjusted binomial}} < 10^{-5}$).

Thus, traditional analyses miss the opportunity to include more linguistic/contextual features, which is somewhat trivial to acknowledge since the traditional analyses simply do not have, or cannot consider, the spreadsheet columns providing such additional information. Traditional analyses also miss the ability to make distinctions even in the columns they might include, namely truly multivariate distinctions: in this case, two separate co-varying collexeme analyses would make the analyst postulate two types {advice, *do*, preposed} and {advice, *do*, postposed} while only the explicit multivariate context reveals that, in comparison, the latter is actually a comparative antitype. Our methodological approach is in line with other recent usage-based studies that have not restricted their attention to just one slot in a construction (e.g., the verb slot), but consider multiple linguistic features at the same time (e.g., Hoffmann *et al.* 2019; Olguín Martínez & Gries 2024).

5.2   Theoretical implications

A number of studies adopting a usage-based approach have shown that linguistic structure is lexically particular "in the sense that grammatical categories and constructions are generally associated with specific lexical items" (Diessel 2019:20). The co-occurrence patterning of lexemes and constructions is functionally motivated (Goldberg 1995:50; Gries & Stefanowitsch 2004a, b), which gives rise to a

joint distribution of lexemes in constructions that are known in the literature as FILLER-SLOT RELATIONS (see Diessel 2019). The fact that the meaning of a construction tends to harmonize with the meanings of the lexical elements that typically occur in it is referred to as the SEMANTIC COHERENCE PRINCIPLE (Goldberg 1995:50). On a most basic level, our work also analyzes the interaction of lexicon and syntax in language use. However, we went beyond this traditional work in that we also explored how other constructional properties can also influence the compatibility of lexemes in particular slots of constructional schemas. For instance, our study has shown that the order of clauses can play an important role in the compatibility of lexemes in the slot of apodoses in counteridentical constructions. The findings of the present study have provided a novel way to explore the following question: how do syntax, semantics, discourse, and lexicon fit together in a unified model of linguistic architecture? The theoretical implications of the present research are problematic for formal linguistic theory, in which the distinction between lexicon and syntax has played an important role.

6.   Conclusions

6.1   Summary

The most important results can be summarized as follows. We showed that counteridenticals containing postposed apodoses prefer to occur with protases with 3rd person, with apodoses with *be able to* and *be*, with present temporality, and in imaginative discourse contexts. From a usage-based perspective, we argued that *be able to* and *be* are verb lemmas used for expressing similarity between a speaker and third person entities in terms of physical qualities and emotional states. Accordingly, their semantics align with imaginative discourse contexts and 3rd person (i.e., similarity between speaker and 3rd person in terms of Y) and the postposed position of apodoses. One antitype of counteridenticals containing postposed apodoses is worth summarizing. Counteridentical constructions containing postposed apodoses dislike 3rd person, future time reference, and the advice function. We argued that, from a usage-based perspective, this dispreferred co-occurrence of features is easy to justify in that advice counteridenticals can only be given in a speaker-addressee-context (2nd person) and occur with preposed apodoses due to their framing potential that is relevant to the organization of ongoing social interaction.

As for counteridenticals containing preposed apodoses, the strongest attraction of preposed apodoses was to the collexemes *touch*, *watch*, *get*, *take*, *worry*, and *do* when protases occur with 2nd person, with future temporal reference, and

in advice contexts. We argued that these verb lemmas occur in advice counteridentical constructions with precautioning semantics. Accordingly, they can be situated within the domain of apprehensives. In this scenario, a speaker either warns about a possible danger that may result in physical damage if the addressee does not carry out the precaution expressed in the apodosis or indicates some situation which the speaker considers to be unpleasant, and which should be avoided. Two antitypes of counteridentical constructions containing postposed apodoses are worth summarizing. First, preposed apodoses do not like to appear with the cluster {2nd person, *be able to*, future time reference, and advice function}. This stems from the fact that *be able to* is not suitable for discourse contexts that have a framing potential that is relevant to the organization of ongoing social interaction (preposed apodoses), i.e., involving a recommendation, suggestion, or advice. Second, preposed apodoses do not like to appear with the cluster {3rd person, *be*, present time reference, and imaginative function}. We argued that this is due to the fact that the verb lemma *be* is used in the expression of similarity between speaker and addressee in terms of physical qualities and emotional states (e.g., *be fast*, *be happy*, *be upset*) in postposed apodoses.

What these results indicate is that a regular collexeme analysis needs to be augmented with more distributional/corpus data/annotation than has been the standard so far. This will enable us to uncover how other constructional properties can also influence the compatibility of lexemes in particular slots of constructional schemas. It is precisely multivariate studies like this that the field needs more of in order to develop a better understanding of the link between syntax, semantics, discourse, and lexicon in language use and building on that, to develop a more holistic framework for linguistic analysis. It is our hope here that the proposed method will be valuable to other usage-based construction grammarians to uncover how speakers entrench various meso-constructions that are characterized by cross-clausal associations.

6.2   Where to go from here?

What are the next steps? In a sense, the next steps follow quite naturally from everything discussed above. First and most obviously, it remains to be analyzed whether counteridentical constructions in which the first-person pronoun does not appear in the protasis (e.g., *in your place/position, I would go*) show similar types and antitypes to the constructions taken into account in the present study. To explore this domain, it would be interesting to apply the method put forward by Olguín Martínez & Gries (2024). This is a new multivariate extension of collostructional analysis that combines distinctive and (co-varying) collexeme analysis via HCFA. It allows the analyst to identify associations not just of one slot to a

construction or one slot to one other slot in one construction, but to include other features to identify (i) which constructions are preferred by which fillers in, now, one or more slots of one construction and (ii) which constructions are preferred by which (combinations) of one or more fillers. For this study, it would be interesting to employ a predictive modeling (e.g., Hampe & Gries 2018) to explore the degree to which the two constructions exhibit distributional patterns that are so distinctive that they are in fact predictive.

The next major kind of follow-up would be to explore how advice counteridentical constructions interact with other discourse factors and sociolinguistic variables by using conversational data. Conversation is a forum that enables interlocuters to calibrate and negotiate experience through language (Levinson & Majid 2014). In the literature, for other types of constructions expressing advice, it has been demonstrated that unsolicited advice-giving actions produce a higher number of instances in which the recipient does not agree with the advice offered than solicited advice-giving actions (Couper-Kuhlen & Thompson 2022:183). Do counteridentical constructions align with this hypothesis? How does this interact with the social roles of the participants, the strength of the stance being taken by the advice-giver (e.g., use of hedges), and the type of advice (e.g., potential physical damage if the advice is not followed)? This remains to be investigated by future studies.

Acknowledgements

References

Arregui, Ana. 2019. Being me, being you: Pronoun puzzles in modal contexts. *Proceedings of Sinn Und Bedeutung* 11. 31–45.

Couper-Kuhlen, Elizabeth & Sandra A. Thompson. 2022. Action ascription and deonticity in everyday advice-giving sequences. In Arnulf Deppermann & Michael Haugh (eds.), *Action ascription in social interaction*, 183–207. Cambridge: CUP.

Declerck, Renaat & Susan Reed. 2001. *Conditionals. A comprehensive empirical analysis*. Berlin: Mouton de Gruyter.

Diessel, Holger. 2008. Iconicity of sequence. A corpus-based analysis of the positioning of temporal adverbial clauses in English. *Cognitive Linguistics* 19. 457–482.

Diessel, Holger. 2019. *The grammar network. How linguistic structure is shaped by language use*. Cambridge: CUP.

Eye, Alexander von. 2002. *Configural frequency analysis: Methods, models, and applications.* Mahway, NJ: Lawrence Erlbaum.

Family, Neiloufar. 2011. Mapping semantic spaces: A constructionist account of the "light verb" in Persian. *Folia Linguistica* 45. 1–30.

Ford, Cecillia E. & Sandra A. Thompson. 1986. Conditionals in discourse: A text-based study from English. In Elizabeth Closs Traugott, Alice Ter Meulen, Judy Snitzer Reilly & Charles A. Ferguson (eds.), *On conditionals*, 353–372. Cambridge: CUP.

Givón, Talmy. 2001. *Syntax: An introduction, Vol. 2.* Amsterdam: John Benjamins.

Goldberg, Adele E. 1995. *Constructions: A construction grammar approach to argument structure.* Chicago, IL: University of Chicago Press.

Goodman, Nelson. 1991. The problem of counterfactual conditionals. In Frank Jackson (ed.), *Conditionals*, 9–27. Oxford: OUP.

Gries, Stefan Th. 2009. *Statistics for linguistics with R: A practical introduction.* Berlin: Mouton de Gruyter.

Gries, Stefan Th. & Anatol Stefanowitsch. 2004a. Extending collostructional analysis: A corpus-based perspective on 'alternations'. *International Journal of Corpus Linguistics* 9. 97–129.

Gries, Stefan Th. & Anatol Stefanowitsch. 2004b. Co-varying collexemes in the *into*-causative. In Michel Achard & Suzanne Kemmer (eds.), *Language, culture, and mind*, 225–236. Stanford, CA: Center for the Study of Language and Information.

Hampe, Beate & Stefan Th. Gries. 2018. Syntax from and for discourse II: More on complex sentences as meso-constructions. In Beate Hampe & Susanne Flach (eds.), *Yearbook of the German Cognitive Linguistics Association*, 115–142. Berlin: Mouton de Gruyter.

Hoffmann, Thomas, Jakob Horsch & Thomas Brunner. 2019. The more data, the better: A usage-based account of the English comparative correlative construction. *Cognitive Linguistics* 30. 1–36.

Kauf, Carina. 2017. *Counterfactuals and (counter-)identity. The identity crisis of "if I were you".* Göttingen: Universität Göttingen MA thesis.

Krauth, Joachim. 1993. *Einführung in die Konfigurationsfrequenzanalyse.* Weinheim: Beltz.

Lakoff, George. 1996. Sorry, I'm not myself today: The metaphor system for conceptualizing the self. In Gilles Fauconnier & Eve Sweetser (eds.), *Spaces, worlds, and grammars*, 91–123. Chicago, IL: University of Chicago Press.

Levinson, Stephen C. & Asifa Majid. 2014. Differential ineffability and the senses. *Mind and Language* 29. 407–427.

Lewis, David. 1973. *Counterfactuals.* Oxford: Blackwell.

Lichtenberk, Frantisek. 1995. Apprehensional epistemics. In Joan Bybee & Suzanne Fleischman (eds.), *Modality in grammar and discourse*, 293–327. Amsterdam: John Benjamins.

Nuyts, Jan. 2015. Subjectivity: Between discourse and conceptualization. *Journal of Pragmatics* 86. 106–110.

Olguín Martínez, Jesús & Stefan Th. Gries. 2024. *If not for — if it weren't/wasn't for* counterfactual constructions: A multivariate extension of collostructional analysis. *Cognitive Semantics* 10. 159–189.

Palmer, Frank R. 2001. *Mood and modality, 2nd edn.* Cambridge: CUP.

Pelletier, Jérôme. 2004. Analogical uses of the first person pronouns: A difficulty in philosophical semantics. *The Journal of Cognitive Science* 5. 139–155.

Preston, Stephanie D. & Frans B. M. de Waal. 2002. Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences* 25. 1–71.

Schmidtke-Bode, Karsten. 2009. *A typology of purpose clauses*. Amsterdam: John Benjamins.

Smith-Dennis, Ellen. 2021. Don't feel obligated, lest it be undesirable: the relationship between prohibitives and apprehensives in Papapana and beyond. *Linguistic Typology* 25. 413–459.

Stefanowitsch, Anatol & Stefan Th. Gries. 2003. Collostructions: Investigating the interaction between words and constructions. *International Journal of Corpus Linguistics* 8. 209–243.

Stefanowitsch, Anatol & Stefan Th. Gries. 2005. Covarying collexemes. *Corpus Linguistics and Linguistic Theory* 1. 1–43.

Sweetser, Eve. 1990. *From etymology to pragmatics: Metaphorical and cultural aspects of semantic structure*. Cambridge: CUP.

Thomas, Guillaume. 2008. Proxy counterfactuals. *Snippets* 18. 17–18.

Van linden, An. 2021. A usage-based approach to counterfactuality: Optionality of the apodosis. *Theoretical Linguistics* 47. 277–286.

Address for correspondence

Jesús Olguín Martínez
University of Hong Kong
Rm 9.30 Run Run Shaw Tower, Centennial Campus, HKU
Hong Kong
olguin@hku.hk
https://orcid.org/0000-0002-4555-4213

Co-author information

Stefan Th. Gries
UC Santa Barbara and JLU Giessen
stgries@linguistics.ucsb.edu

Publication history