Extraction from NP, Frequency, and Minimalist Gradient Harmonic Grammar

Gereon Müller & Johannes Englisch (Universität Leipzig) March 12, 2020

Abstract

Extraction of a PP from an NP in German is possible only if the head noun and the governing verb together form a natural predicate (Müller (1995), Sauerland (1995), Schmellentin (2006)). We show that this corresponds to collocational frequency of the verb-noun combinations in corpora, based on the metric of ΔP (Gries (2013)). From this we conclude that frequency should be conceived of as a language-external grammatical building block that can directly interact with language-internal grammatical building blocks (like triggers for movement and economy constraints blocking movement) in excitatory and inhibitory ways. Integrating frequency directly into the syntax is not an option in most current grammatical theories. However, things are different in *Gradient Harmonic Grammar* (Smolensky & Goldrick (2016)), a version of Optimality Theory where linguistic objects of various kinds can be assigned *strength* in the form of numerical values. We show that by combining a *Minimalist* approach to syntactic derivations (Chomsky (2001)) with a Gradient Harmonic Grammar approach of constraint evaluation, the role of frequency in licensing extraction from PP in German can be integrated straightforwardly, the only additional prerequisite being that (verb-noun) *dependencies* (Manzini (1995), O'Grady (1998), Osborne, Putnam & Groß (2013)), Bowers (2017), Bruening (2018)) qualify as linguistic objects that can be assigned strength (based on their frequency).

1. Extraction from NP

It has often been noted that extraction from NP in German is subject both to structural and to lexical restrictions; cf. Fanselow (1987, ch. 2), Grewendorf (1989, ch. 2.8), Webelhuth (1988; 1992), Müller (1991; 1995; 2011), Sauerland (1995), De Kuthy & Meurers (2001), Schmellentin (2006), Ott (2011), and Frey (2015); also see Cattell (1976), Bach & Horn (1976), Chomsky (1977), Davies & Dubinsky (2003) and Koster (1987, ch. 4) for English and Dutch, respectively.[1] The examples in (1) illustrate extraction from NP in German. As shown by (1-ab) and (1-cd), wh-movement and scrambling can bring about extraction from NP; more

[1] Two remarks. First, throughout this paper, we assume that nominal projections in German are NPs (with DPs as specifiers) rather than DPs (with NPs as complements); see Bruening (2009; 2020), Georgi & Müller (2010), and Bruening et al. (2018), among others. As a matter of fact, the dependence of extraction from nominal projections on a close relation of V and N that is at the heart of the present study can be viewed as a further argument in support of the NP-over-DP hypothesis. That said, by relaxing locality requirements for selection in head-head dependencies, the main claim of the present paper – viz., that collocational frequency can be assumed to directly play a role in licensing extraction – could in principle also be formulated in a DP-over-NP approach to nominal projections in German. Second, some kinds of extraction from NP in German are subject only to structural restrictions (the position of the NP in the clausal spine), not to lexical ones; e.g., this holds for so-called *was-für* split constructions (see Müller (1995) for a characterization of this asymmetry). For the purposes of the present paper, we can leave open the question of why *was für* split does not require a close relation of V and N.

generally, the operation is not confined to specific movement types. Furthermore, the operation can involve either complete PP complements of N, as in (1-ac), or R-pronouns that act as complements of the P heads of complements of N, as in (1-bd); the latter option is restricted to varieties of German that allow postposition stranding more generally (and, in the examples here, with a bare vocalic onset of the preposition in particular; see Riemsdijk (1978), Trissler (1993), and Hein & Barnickel (2018), among others).

(1) a. [$_{\text{PP}_1}$ Worüber] hat der Fritz [$_{\text{NP}}$ ein Buch t$_1$] gelesen ?
about.what has the Fritz$_{nom}$ a book$_{acc}$ read

b. [$_{\text{DP}_1}$ Wo] hat die Maria [$_{\text{NP}}$ ein Buch [$_{\text{PP}}$ über t$_1$]] gelesen ?
what has the Maria$_{nom}$ a book$_{acc}$ about read

c. dass [$_{\text{PP}_1}$ darüber] keiner je [$_{\text{NP}}$ ein Buch t$_1$] gelesen hat
that about that no-one$_{nom}$ ever a book$_{acc}$ read has

d. dass [$_{\text{DP}_1}$ da] keiner je [$_{\text{NP}}$ ein Buch [$_{\text{PP}}$ t$_1$ über]] gelesen hat
that that no-one$_{nom}$ ever a book$_{acc}$ about read has

Among the structural factors restricting the operation we take to be the following. First, extraction from NP is not possible with external arguments (of transitive or unergative verbs); cf. (2).

(2) *[$_{\text{PP}_1}$ Worüber] hat [$_{\text{NP}}$ ein Buch t$_1$] den Fritz beeindruckt ?
about.what has a book$_{nom}$ the Fritz$_{acc}$ impressed

Next, extraction from NP cannot take place with indirect objects bearing dative case (cf. (3-a)), even if the verb as such allows extraction from NP (cf. (3-b), where extraction from the direct object occurs in a ditransitive, dative-accusative environment).

(3) a. *[$_{\text{PP}_1}$ Worüber] hat man [$_{\text{NP}}$ einem Buch t$_1$] einen Preis gegeben ?
about what has one$_{nom}$ a book$_{dat}$ an award$_{acc}$ given

b. [$_{\text{PP}_1}$ Worüber] hat man der Maria [$_{\text{NP}}$ ein Buch t$_1$] gegeben ?
about what has one$_{nom}$ the Maria$_{dat}$ a book$_{acc}$ given

Third, extraction from a definite NP typically yields degraded results; this specificity effect (cf. Mahajan (1992), Webelhuth (1992)) is shown in (4), which forms a minimal pair with the non-specific example in (1-a).

(4) ?*[$_{\text{PP}_1}$ Worüber] hat der Fritz [$_{\text{NP}}$ das Buch t$_1$] gelesen ?
about what has the Fritz$_{nom}$ the book$_{acc}$ read

A fourth observation is that extraction from NP is blocked when there is a possessor NP present (either pre-nominally or post-nominally); see (5).

(5) *[$_{\text{PP}_1}$ Worüber] hat die Maria [$_{\text{NP}}$ Fritzens Buch t$_1$] gelesen ?
about what has the Maria$_{nom}$ Fritz$_{gen}$ book$_{acc}$ read

Finally, freezing effects occur if a direct object which as such licenses extraction undergoes movement itself. Thus, (6) illustrates that an NP blocks extraction if it is scrambled; compare (6-a) with (6-b).[2]

(6) a. *[$_{\text{PP}_1}$ Worüber] hat [$_{\text{NP}_2}$ ein Buch t$_1$] keiner t$_2$ gelesen ?
about what has a book$_{acc}$ no-one$_{nom}$ read

b. [$_{\text{PP}_1}$ Worüber] hat keiner [$_{\text{NP}_2}$ ein Buch t$_1$] gelesen ?
about.what has no-one$_{nom}$ a book$_{acc}$ read

All of these structural restrictions on extraction from NP can be derived without too much ado under current approaches to movement, based on (whatever derives) the Condition on Extraction Domain (Huang (1982), Chomsky (1986)) and the Minimal Link Condition (Chomsky (2001; 2008)); see, e.g, Müller (2011) for an account of these phenomena that relies on Chomsky's (2001) Phase Impenetrability Condition (PIC).[3]

In addition to these structural factors, extraction from NP in German is conditioned by lexical factors. Thus, whereas a verb like *lesen* ('read') in (1-a) (repeated here as (7-a)) permits extraction from the NP headed by *Buch* ('book'), a verb like *stehlen* ('steal') does not, in an identical environment (see (7-b)). Note that syntactically, the two verbs otherwise behave the same (they take an internal theme argument as a direct object and an external agent argument as a subject, they assign accusative to the direct object, etc.). What is more, as observed by Sauerland (1995), not only is nature of the verb relevant: By keeping the verb identical and modulating the head noun of the object, extraction can also become impossible; see (7-c), where *Verlautbarung* ('official statement') replaces *Buch* ('book') in the presence of *lesen* ('read'). As one might expect, a combination as in (7-d) will also block extraction from NP: Here *Verlautbarung* is the head noun and *stehlen* is the governing verb.

[2] Also note that scrambling of the indefinite NP to a position in front of the external argument NP *keiner* ('no-one$_{nom}$'), although slightly marked and dependent on appropriate contexts and intonation, is well formed as such when there is no concurrent extraction from NP; see (i).

(i) dass [$_{\text{NP}_2}$ ein Buch über dieses Thema] keiner t$_2$ gelesen hat
that a book$_{acc}$ about this topic no-one$_{nom}$ read has

[3] Some of the structural restrictions on extraction from NP in German have sometimes been disputed. Thus, Haider (1983; 1993) and Diesing (1992) have argued that subject DPs can also be transparent for extraction in German. Howoever, Fanselow (2001, 422) has shown that the vast majority of what at first sight might look like counter-examples to the claim that subject NP are islands for extraction of (and out of) PPs in German involve passive or unaccusative constructions with the nominative DP in situ, in a complement (i.e., object) position, as in (i-ab).

(i) a. [$_{\text{PP}_1}$ Über wen] wurde [$_{\text{DP}}$ ein Buch t$_1$] gelesen ?
about whom was a book$_{nom}$ read

b. [$_{\text{PP}_1}$ Über wen] ist [$_{\text{DP}}$ ein Buch t$_1$] erschienen ?
about whom is a book$_{nom}$ appeared

Other apparent counter-examples involve PPs headed by *von* ('of'), as in (iii-a), or *zu* ('to'), as in (iii-b) (see De Kuthy & Meurers (2001, 149) and Haider (1993, 172-173), among others). For these, an analysis that does not

(7) a. [$_{PP_1}$ Worüber] hat der Fritz [$_{NP}$ ein Buch t_1] gelesen ?
about.what has the Fritz$_{nom}$ a book$_{acc}$ read

b. *[$_{PP_1}$ Worüber] hat der Fritz [$_{NP}$ ein Buch t_1] gestohlen ?
about.what has the Fritz$_{nom}$ a book$_{acc}$ stolen

c. ?*[$_{PP_1}$ Worüber] hat der Fritz [$_{NP}$ eine Verlautbarung t_1] gelesen ?
about.what has the Fritz$_{nom}$ an official.statement$_{acc}$ read

d. *[$_{PP_1}$ Worüber] hat der Fritz [$_{NP}$ eine Verlautbarung t_1] gestohlen ?
about.what has the Fritz$_{nom}$ an official.statement$_{acc}$ stolen

The conclusion that suggests itself in view of this kind of evidence is that for extraction from NP of a PP complement (or an R-pronoun contained in it) to be legitimate in German, V and N must enter a tight relationship; they must form a *natural predicate*, i.e., a dependency of two lexical items that qualifies as entrenched.

It is not a priori clear how this condition can be implemented in grammatical theory. Following Bach & Horn's (1976) proposal for English, Fanselow (1987) assumes that extraction from NP is in fact never possible in German; rather, data of the kind in (7-a) are the result of a pre-syntactic *reanalysis* rule that makes it possible for the verb to take not just NP, but also PP directly as arguments, so that PP does not have to leave NP in (7-a) in the first place. Whereas a reanalysis approach along these lines has sometimes been been adopted by subsequent studies (cf., e.g., De Kuthy (2001); De Kuthy & Meurers (2001)), severe problems have been pointed out for it that, in our view, make such an analysis untenable (see Webelhuth (1988), Fanselow (1991), Müller (1998), and Schmellentin (2006), among others). For one thing, in the absence of a theory of general restrictions on reanalysis rules, it is completely unclear why reanalysis cannot involve a verb and agent (subject) or goal (indirect object) arguments; recall (2), (3-a). Next, on this view, it is a mystery why specificity and possessor intervention effects should arise if there is no extraction from NP in the first place; see (4), (5). Next, if PP does not have to undergo extraction from NP in the well-formed examples discussed so far, how can it be that NP scrambling creates a typical freezing effect (as in (6-a) vs. (6-b))?

Now, it is known that verbs like *lesen* ('read') in (7-a), in contrast to verbs like *stehlen* ('steal') in (7-b), may occur in constructions in which the PP is present but the NP is either completely absent or realized only as a pronoun. This is generally taken to be the strongest argument in support of the base-generation approach to extraction from NP; see (8-a) vs. (8-b).

---

involve actual extraction seems systematically available. For instance, *von-* ('by') phrases are known to often involve external generation of an optional argument instead of extraction (see Koster (1987, 196f.), Cinque (1990, 47), Sternefeld (1991, 121), Müller (1995, 397f.), Barbiers (2002, 54), and Gallego (2007, 349), among others).

(ii) a. [$_{PP_1}$ Von den Studenten] haben [$_{NP}$ viele (t_1)] die Prüfung nicht geschafft
of the students have many the exam not made

b. ?[$_{PP_1}$ Zu diesem Problem] haben [$_{NP}$ einige Briefe (t_1)] den Sender erreicht
to/concerning this problem have several letters$_{nom}$ the station reached

(8) a. dass Fritz (?es) [$_{PP}$ über die Liebe] gelesen hat
that Fritz it about the love read has

b. *dass Fritz (es) [$_{PP}$ über die Liebe] geklaut hat
that Fritz it about the love stolen has

However, verbs like *geben* ('give') in German behave like *lesen* in that they permit extraction from a direct object NP (cf. (3-b)), but behave like *stehlen* ('steal') in that they do not allow the NP to be pronominal or dropped (cf. (9-a)). What is more, as shown in (7-c), *lesen* ('read') does not permit extraction if the head noun of its complement is *Verlautbarung* ('official.statement'), but NP can be pronominal or zero in this context; see (9-b). Thus, the correlation breaks down, in both directions (there is the option of extraction without the option of pronominal/zero realization of NP, and there is the option of pronominal/zero realization of NP without the option of extraction); and with it goes the argument for reanalysis.[4]

(9) a. *dass man (es) der Maria [$_{PP}$ über die Liebe] gegeben hat
that one$_{nom}$ it$_{acc}$ the Maria$_{dat}$ about the love given has

b. dass der Fritz (sie) [$_{PP}$ über die Liebe] gelesen hat
that the Fritz$_{nom}$ she$_{acc}$ about the love read has

To conclude, reanalysis as a tool to account for extraction from NP is problematic from an empirical point of view. Furthermore, as noted above, there is no theory of what a reanalysis rule can and cannot look like; more generally, the concept emerges as dubious from a conceptual point of view, too (see, e.g., Baltin & Postal (1996, 135ff)).

At this point, two basic questions need to be addressed as regards the influence of lexical factors on extraction from NP. The first question is how it can be determined whether a V and an N can form a natural predicate; i.e., how this lexical factor can be measured. And the second question is how this information then licenses or blocks the grammatical process of extraction, i.e., how the lexical factor, once its nature is determined, can interact with the building blocks of grammar that are involved in syntactic movement.[5] In a nutshell, the answers we will give are that the concept of a natural predicate corresponds to colloca-

[4] We cannot offer a full-fledged account of the pronominal and zero forms showing up here. Assuming a post-syntactic approach to morphological realization (cf. Halle & Marantz (1993), Marantz (1995)) would provide a reasonably straightforward analysis (on this view, N and its specifier(s) could in principle be realized by lexical categories, as in *ein Buch*, or by a pronoun, as in *es*, or not at all (Ø)). However, given the observation that the specific choice of lexical item for V and N determines the felicity of the syntactic operation of extraction, this does not in fact look like a viable option. For present purposes, we will simply assume that the zero and pronominal realizations are pre-syntactically determined versions of N which can take a PP complement but do not permit any pre-nominal items (like determiners or adjectives); such NPs may then qualify as unstable structures.

[5] Incidentally, it is worth noting that even if pre-syntactic reanalysis were to provide the correct approach to extraction from NP, essentially the same two questions would arise that arise under the view that extraction from NP is possible in principle. The first question would be identical, and the second question would then be how fixing the first factor can influence the application of the reanalysis rule.

tional frequency, which can be encoded as a numerical value for V-N dependencies (section 2); and that an approach to syntax that combines Minimalist derivations with constraint interaction in a Gradient Harmonic Grammar approach makes it possibe to implement the lexical factor, by letting the numerical values capturing different collocational strengths of V-N dependencies interact with constraints that trigger and block extraction (section 3).

2. Frequency

2.1. ΔP

In what follows, we will pursue the hypothesis that frequency is the decisive factor in establishing a natural predicate, i.e., an entrenched V-N dependency, in the cases of extraction from NP that we are interested in. A basic premise is that the absolute frequencies of individual lexical items in corpora will not be particularly informative in this context, and that the same goes for the absolute frequencies of V-N collocations. Rather, what is needed is a more fine-grained approach to frequency that is based on how well the two lexical items in a V-N dependency predict each other. One such measure that has been proposed is *collostructional strength* (see Stefanowitsch (2009), Gries & Stefanowitsch (2004), and Gries et al. (2005)). More recently, Gries (2013) has suggested to employ the measure of ΔP, and it is this concept that we will make use of in what follows. $\Delta P_{X|Y}$ measures how well the presence of some item Y predicts the presence or absence of some other item X. ΔP is defined as in (10).

(10) ΔP (Gries (2013, 143)):
ΔP = *p(outcome|cue = present) – p(outcome|cue = absent)*

Here, *p(X|Y = present)* captures the probability of the outcome X in the presence of the cue Y; *p(X|Y = absent)* is the probability of the outcome X in the absence of the cue Y; and to determine $\Delta P_{X|Y}$, the latter is subtracted from the former. The values of ΔP range from -1.0 to 1.0; they are interpreted as follows:

- $\Delta P_{X|Y}$ approaching 1.0: Y is a good cue for the presence of X
- $\Delta P_{X|Y}$ approaching -1.0: Y is a good cue for the absence of X
- $\Delta P_{X|Y}$ approaching 0.0: Y is not a good cue for the presence or absence of X

Note that this relationship is asymmetric. An element predicting another element well is not necessarily well-predicted by that element. This means that for every pair of X and Y, there are two values $\Delta P_{X|Y}$ and $\Delta P_{Y|X}$. As an illustration, let us look at how ΔPs are determined for a V-N dependency involving *kaufen* ('buy') and *Buch* ('book') on the basis of the frequencies of the co-occurrences. To calculate $\Delta P_{X|Y}$, we first search the corpus for the number of all cases where X and Y co-occur, where only one of the elements occurs, and where none of the elements occur. (11) shows such a co-occurrence table for the pair *Buch kaufen* 'book buy'.[6]

(11) *Co-occurences of* Buch *'book' and* kaufen *'buy'*

Buch kaufen	V present	V absent	totals
N present	142	25696	25838
N absent	8370	5785365	5793735
totals	8512	5811061	5819573

This kind of information can be used to calculate $\Delta P_{X|Y}$ by taking the difference between the probability of X given the presence of Y and the probability of X given the absence of Y. Suppose that X = *Buch* and and Y = *kaufen*. $\Delta \mathrm{P}_{X|Y}$ (= $\Delta \mathrm{P}_{\mathrm{Buch|kaufen}}$) is then determined as shown in (12); it shows how well *kaufen* predicts *Buch* in the corpus.

(12)
$$\begin{aligned}\Delta P_{\text{Buch|kaufen}} &= p(\text{Buch}|\text{kaufen present}) - p(\text{Buch}|\text{kaufen absent})\\ &= \frac{\text{Buch and kaufen present}}{\text{kaufen present}} - \frac{\text{Buch present and kaufen absent}}{\text{kaufen absent}}\\ &= \frac{142}{8512} - \frac{25696}{5811061}\\ &\approx 0.0123\end{aligned}$$

In the same way, $\Delta P_{Y|X}$ (= $\Delta P_{\text{kaufen|Buch}}$) based on the data in (11) is calculated as shown in (13). The resulting value indicates how well *Buch* predicts *kaufen*.

(13)
$$\begin{aligned}\Delta P_{\text{kaufen|Buch}} &= p(\text{kaufen}|\text{Buch present}) - p(\text{kaufen}|\text{Buch absent})\\ &= \frac{\text{kaufen and Buch present}}{\text{Buch present}} - \frac{\text{kaufen present and Buch absent}}{\text{Buch absent}}\\ &= \frac{142}{25838} - \frac{8370}{5793735}\\ &\approx 0.0041\end{aligned}$$

By comparing the two ΔPs, it becomes evident that *kaufen* is a somewhat better predictor for *Buch* than *Buch* is for *kaufen*: The likelihood of a buying event involving books (rather than, say, bikes or guitars) is greater (ΔP = 0.0123) than the likelihood that a book is involved in a buying event (rather than, say, a reading or burning event, or some other scenario in which books may show up; ΔP = 0.0041).[7]

[6] The numbers here are actual numbers based on our corpus study; see the next section.

[7] As a side remark, note that this approach of ΔP determination based on corpus frequencies can run into a technical problem if there are zero counts of some item in the corpus. As an example, consider the data in (i).

(i) a. Der Fritz hat [$_{\text{NP}}$ ein Buch über Tiere] geliket
the Fritz$_{nom}$ has a book$_{acc}$ about animals a.like.given

b. *Worüber$_1$ hat der Fritz [$_{\text{NP}}$ ein Buch t$_1$] geliket?
about.what has the Fritz$_{nom}$ a book a.like.given

The verb *liken* ('to give a like (on social media)') is a loanword from English in German that did not exist in the 20th century and is therefore not attested in the corpus that we base our analysis on (see the next subsection). In line with this, we obtain results for the V-N dependency consisting of *liken* and *Buch* where $\Delta P_{\text{liken|Buch}}$ is

2.2. Corpus

The data in our survey come from the core corpus of *Digitales Wörterbuch der deutschen Sprache* (DWDS; see Geyken (2007)). The DWDS is a freely searchable corpus consisting of about 5.8 m sentences in the German language. It contains a balanced mix of fictional, scientific, functional, and newspaper texts from the 20th century.

The list in (14) shows the queries used to elicit the counts for nouns, verbs, and noun-verb pairs. In a perfect world, we would like to query the corpus for every instance where a given noun is the direct object of a given verb (recall that this is the only environment in which extraction from NP can be possible, given our characterization of the empirical evidence in the previous section). However, while the corpus is lemmatised and tagged for part-of-speech, it does not encode dependencies. Because of this the queries can only ever be approximations.

(14) a. Query: `Buch with $p=NN`
Searches for the lemma *Buch* with the part-of-speech tag NN (common nouns)
b. Query: `kaufen with $p=VV*`
Searches for the lemma *kaufen* with a part-of-speech tag starting with VV (verbs)
c. Query: `near(Buch with $p=NN, kaufen with $p=VV*, 3)`
Searches for a sentence with the noun *Buch* and the verb *kaufen*, with zero to three tokens between them.

The query in (14-c) attempts to find noun-verb pairs by looking for sentences where the noun and the verb are close to each other. This avoids false positives as in (15-a) (where *Buch* ('book') and *gekauft* ('bought') are clause-mates in a VP coordination construction, but *Buch* is the (head of the) object of *gelesen* ('read'), not of *gekauft*). However, it also introduces false negatives as in (15-b), where *Buch* is the (head of the) object of *gelesen*, but is separated from it by more than three items as a consequence of having undergone topicalization to the clause-initial ('Vorfeld') position.

(15) a. Fritz hat ein *Buch* gelesen und ein Lesezeichen *gekauft*.
Fritz has a book read and a bookmark bought
'Fritz read a book and bought a bookmark.'
b. Das *Buch* hat der Fritz in der Innenstadt *gekauft*.
the book has the Fritz in the city.centre bought
'Fritz bought the book in the city centre.'

Cases like (15-b) can only pose a potential problem if there is reason to assume that object topicalization also (i.e., like extraction from NP) shows asymmetries depending on how

0 whereas $\Delta P_{\mathrm{Buch|liken}}$ is in fact not solvable because division by zero is undefined. In what follows, we will abstract away from this technical issue which does not arise in practice (if we want to determine whether extraction from NP is possible with a certain kind of verb, that verb must exist, however rare its occurrence may be).

close the relation between the verb and the object's head noun is, such that, e.g., an object headed by *Buch* ('book') tends to undergo topicalization more often, or more easily (or, in fact, less often, or less easily) in the presence of *lesen* ('read') than in the presence of *stehlen* ('steal'). We are not aware of any claims in the literature that would go in this direction, and will assume, here and henceforth, that there is no such effect. Thus, false negatives like (15-b) generated by object movement can be ignored, assuming that they affect all kinds of V-N dependencies in the same way.[8]

2.3. *Results*

We have determined both ΔP values for every V-N pair where N is a noun in in (16-a) and V is a verb in (16-b). This results in high-frequency collocations like *Buch lesen* ('book read'), combinations of low-frequency pairs where this intuitively seems to be 'the noun's fault', like *Verlautbarung lesen* ('official.statement read'), and combinations where it is the verb that is responsible for the low frequency, as in *Buch werfen* ('book throw').

(16) a. Nouns
Bericht 'report', *Buch* 'book', *Geschichte* 'story/history', *Roman* 'novel', *Verlautbarung* 'official statement'

b. Verbs
aufschlagen 'open (book)', *kaufen* 'buy', *klauen* 'steal (coll.)', *lesen* 'read', *öffnen* 'open', *schreiben* 'write', *stehlen* 'steal', *verfassen* 'write (book)', *verkaufen* 'sell', *vorlesen* 'read (to sb.)', *weglegen* 'put away', *werfen* 'throw'

As shown in (18), the ΔPs for *Buch lesen* are both higher than the the ΔPs for *Buch stehlen.*

(17) *ΔPs for two V-N pairs*:

N	V	extraction from NP	$\Delta P_{V\|N}$	$\Delta P_{N\|V}$
Buch 'book'	lesen 'read'	yes	0.027000422	0.035377353
Buch 'book'	stehlen 'steal'	no	0.000105424	0.001296759

This is in full accordance with the fact that the V-N dependency *Buch lesen* permits extraction from the NP whereas the V-N dependency *Buch stehlen* does not; recall the examples in (7-a) and (7-b) above. As shown by the ΔPs for some other V-N combinations in (18), this result can be generalized: The higher a ΔP is, the more likely it is that extraction is possible.

[8] The same conclusion can be drawn for cases where verb-second movement leads to a larger distance between V and N than the one covered by the query.

(18) *ΔPs for more V-N pairs*:

N	V	extr./NP	$\Delta P_{V\|N}$	$\Delta P_{N\|V}$
Buch 'book'	schreiben 'write'	yes	0.022816910	0.016380559
Buch 'book'	kaufen 'buy'	yes	0.004051117	0.012260419
Bericht 'report'	schreiben 'write'	yes	0.001723002	0.000636521
Buch 'book'	weglegen 'put away'	no	0.000332618	0.085561626
Buch 'book'	öffnen 'open'	no	-0.001213743	-0.002659810
Bericht 'report'	werfen 'throw'	no	-0.002749089	-0.002112361
Verlautbarung 'off.st.'	lesen 'read'	no	-0.000209264	-0.000005488

These data also shed some light on which ΔP value may be most relevant for establishing the strength of a V-N dependency (and, consequently, for determining the option of extraction from NP). A priori, three options suggest themselves: $\Delta P_{V|N}$, $\Delta P_{N|V}$, and the arithmetic mean of these two values. Closer inspection reveals that $\Delta P_{N|V}$ is not fully reliable. On the one hand, there are cases like *Buch weglegen* ('book put.away') where $\Delta P_{N|V}$ is fairly high (i.e., *weglegen* ('put.away') is a reasonably good predictor for the presence *Buch* ('book')), but extraction is not possible in this environment (cf. **Worüber hat der Fritz ein Buch weggelegt?*, 'about.what has the Fritz$_{nom}$ a book$_{acc}$ put.away'). On the other hand, there are also cases like *Bericht schreiben* ('report write') where $\Delta P_{N|V}$ is quite low (i.e., *schreiben* ('write') is not a good predictor for the presence of *Bericht* ('report')), but extraction is easily possible (cf. *Worüber hat der Fritz einen Bericht geschrieben?*, 'about.what has the Fritz$_{nom}$ a report$_{acc}$ filed'). In contrast, $\Delta P_{V|N}$ makes the right predictions in these cases: *Bericht* ('report') is a good predictor for *schreiben* ('write'), and *Buch* ('book') is not a good predictor for *weglegen* ('put.away'). This leaves $\Delta P_{V|N}$ and the arithmetic mean of the values as the remaining options. In what follows, we will settle for $\Delta P_{V|N}$ alone. Note that this introduces an asymmetry: Whether a V-N dependency qualifies as a natural predicte or not depends on how well the noun can predict the verb.

2.4. Scaling

In the next section, we will implement the frequency-based approach to extraction from NP in German in a version of Gradient Harmonic Grammar (see Smolensky & Goldrick (2016)). Standardly, numerical strength values assigned to linguistic objects in this grammatical theory are taken to be within the interval [0,1].[9] To simplify exposition, we will therefore scale up numerical values of the type found for ΔP in (17) and (18), so that they end up squarely in the [0,1] interval, making use of the whole range (negative ΔP values will uniformly be treated as 0.) There are several options that suggest themselves; in what follows, we focus on two of them. The first option consists in dividing all values by the maximal ΔP value we have found (i.e., that of *Buch lesen*). This method is faithful to the original distances between values. The second option is slightly more involved. We multiply the ΔP values by

[9] Note, though, that this is stricly speaking a convention, and not a technical limitation. Gradient Harmonic Grammar as such does not explicitly rule out negative, or large, numbers.

10000, leaving the relative distances between the different V-N dependencies intact so far, and then apply a logarithm function to base 1000 to the resulting numbers. This does not leave the original relations between the values intact, but it yields coarser values with clear distances of a similar type among them (even in cases of radically diverging original ΔP values), which can be more easily integrated with simple values assigned to the constraints of the grammatical model as they are standardly postulated. The two scalings produce the strength assignments in (19) for the V-N dependencies in (17) and (18).

(19) *Strength assignments for V-N dependencies*:

N	V	$\Delta P_{V\|N}$	/0.027000422	x10^4; $\log_{10^3}$
Buch	lesen	0.027000422	1.0	0.81
Buch	schreiben	0.022816910	0.845057533	0.79
Buch	kaufen	0.004051117	0.150039025	0.54
Bericht	schreiben	0.001723002	0.063814632	0.41
Buch	weglegen	0.000332618	0.012318993	0.17
Buch	stehlen	0.000105424	0.003904532	0.01
Buch	öffnen	-0.001213743	0	0
Bericht	werfen	-0.002749089	0	0
Verlautbarung	lesen	-0.000209264	0	0

We will adopt the values in the final column in what follows. On the one hand, as just noted, this simplifies exposition. On the other hand, it reflects the assumption that differences between ΔPs in low-frequency domains seem to be more relevant than differences between ΔPs in high-frequency areas.

3. Minimalist Gradient Harmonic Grammar

3.1. *Gradient Harmonic Grammar*

Harmonic Grammar (see Smolensky & Legendre (2006), Pater (2016)) is a version of Optimality Theory (see Prince & Smolensky (1993)) that abandons the strict domination property (according to which no number of violations of lower-ranked constraints can outweigh a single violation of a higher-ranked constraint) and replaces harmony evaluation by constraint ranking with harmony evaluation based on weight assignment to constraints. The central concept of harmony is defined in (20) (see Pater (2009)).

(20) *Harmony*:

$$H = \sum_{k=1}^{K} s_k w_k \qquad w_k = \text{weight of a constraint; } s_k = \text{violation score of a candidate}$$

According to (20), the weight of a constraint is multiplied with the violation score of a candidate for that constraint, and all the resulting numbers are added up, thereby determining the harmony score of a candidate. An output qualifies as optimal if it is the candidate with maximal harmony in its candidate set; i.e., if it has the highest harmony value.

Gradient Harmonic Grammar (see Smolensky & Goldrick (2016)), in turn, is an extension of Harmonic Grammar where it is not just the constraints that are given weights; rather, symbols in linguistic representations are also assigned weights (between 0 and 1). This gives rise to a very straightforward way of associating *strength* with linguistic objects. So far, most of the work on gradient harmonic grammar has been in phonology; but cf. Smolensky (2017), **?**, Lee (2018), Müller (2019), and **?** for recent applications in syntax.[10]

3.2. Minimalist Derivations

We adopt a minimalist setting (cf. Chomsky (2001)), according to which syntactic structure is created incrementally by external and internal Merge operations, where the former are responsible for basic structure-building and the latter bring about structure-building by movement. We assume that syntactic movement is restricted by the inviolable Phase Impenetrability Condition (PIC; cf. Chomsky (2001; 2008))) in (21).[11]

(21) *Phase Impenetrability Condition* (PIC; inviolable):
The domain of a head X of a phase XP is not accessible to operations outside XP; only X and its edge are accessible to such operations.

This implies that movement must take place successive-cyclically, via intermediate edge domains (i.e., specifiers) of phases, where the clausal spine is composed of CP, TP, vP, and VP, of which CP and vP qualify as phases. (We follow Chomsky in assuming that NP/DP is not a phase.) Next, suppose that all Merge operations, including movement steps to intermediate phase edges, are triggered by designated features (cf. Chomsky (1995; 2001), Pesetsky & Torrego (2006), Urk (2015), Collins & Stabler (2016) and Georgi (2017)); this can be enforced by the MERGE CONDITION (MC) in (22) (see Heck & Müller (2013) for the [•F•] notation for features that trigger external or internal Merge), which we assume to be a violable, weighted constraint (in contrast to the PIC).

(22) MERGE CONDITION (MC: violable, weighted):
For all features [•F•], [•F•] triggers Merge of an XP with a matching [F].

Next, there is a counteracting constraint that prohibits structure-building; for present purposes, it can be assumed that this role is played by the ECONOMY CONDITION (EC) in (23)

[10] As a matter of fact, Squishy Grammar as developed in Ross (1973a;b; 1975) is an immediate predecessor of Gradient Harmonic Grammar in syntax. It is interesting to note that even though Squishy Grammar is widely regarded as having been refuted once and for all (see Gazdar & Klein (1978) and Newmeyer (1986; 2000)), closer scrutiny reveals that very few actual counter-arguments against this approach have been presented. However, a detailed reconsideration of the original counter-arguments, while certainly worthwile, is beyond the scope of the present paper.

[11] The status of the PIC as an inviolable constraint is arguably to be expected if one assumes that this constraint is really a theorem derived from basic assumptions about the nature of cyclic spell-out, as suggested by Chomsky.

(see Grimshaw (1997), Legendre et al. (2006); also see Grimshaw (2006) for an attempt at a yet more principled approach). Like MC, EC is violable, and associated with a weight.

(23) ECONOMY CONDITION (EC: violable, weighted):
Merge is prohibited.

Given this state of affairs, for now it looks as though the relative weights of MC and EC decide on whether Merge can apply or not. In a pure Harmonic Grammar approach, this may indeed be true (abstracting away from the potential influence of other constraints for the time being). However, in Gradient Harmonic Grammar, things are somewhat more flexible since the varying strength of the [•F•] features that MC is formulated as a restriction for lead to different degrees of violation of this constraint. A [•F•] feature with a weight of 0.2 will trigger a less severe violation of MC in an output where movement does not take place than a [•F•] feature with a weight of 0.6, and this may distinguish between a violation of MC (in a candidate that does not carry out movement) that is optimal and one that is not. As shown in Müller (2019), asymmetries between different kinds of Merge operations – in particular, between different types of movement – can be derived in such an MC/EC-based approach by postulating different weights (like 0.2 vs. 0.6) of the individual [•F•] features that trigger the operations. With stronger features, an MC violation may become fatal that may be tolerable with weaker features; stronger features may thus ensure that structure-building (or movement) takes place where weaker features do not. This way, it can be derived that, e.g., wh-movement (with a strong trigger [•wh•] on C) can leave a CP in German whereas scrambling (with a weak trigger [•Σ•] on v or V) cannot do so. That said, as shown in section 1, extraction from NP in German does not distinguish between wh-movement and scrambling (or, for that matter, topicalization, relativization, or others movement types that exist in German); cf. Webelhuth (1992), Müller (1995)). For this reason, to keep things simple, we will postulate in what follows that a violation of MC is always of strength -1.0, independently of which movement type is involved.

Against this background, two questions need to be answered to provide an account of extraction from NP in German. First, how does optimization of Merge operations proceed technically? And second, how can the (frequency-based) weights assigned to V-N dependencies be integrated as a factor that may enable or block extraction from NP in the presence of MC and EC? We address the two issues in turn.

3.3. Optimization

There are two general possibilities to model the interaction of minimalist derivations and harmony evaluation. A first option is that all syntactic operations (which, by assumption, take place in the *Gen* component of the grammar) precede a single, parallel step of harmony evaluation (*H-Eval*). This then qualifies as a standard case of harmonic parallelism (see Prince & Smolensky (2004)), and it has been explicitly pursued by, e.g., Broekhuis (2006) and Broekhuis & Woolford (2013). Another option is that Merge operations (GEN) and harmony evaluation (H-EVAL) alternate constantly. On this view, syntactic operations and selection of the most harmonic (optimal) output are intertwined. This model is an instance of *harmonic serialism* (see Prince & Smolensky (2004)). It has been adopted in, e.g., Heck & Müller (2013) and Murphy (2017) (also see McCarthy (2010) and contributions in McCarthy & Pater (2016) for some applications in phonology). In what follows, we adopt an approach based on harmonic serialism. Harmonic serialism in syntax can be viewed as a procedure that is actually little more than a reasonably precise specification of standard minimalist approaches that incorporate a concept of the *best next step* at any given stage of the derivation (see, e.g., Chomsky (1995; 2001) on Merge over Move). The mechanics of harmonic serialism are laid out in (24).

(24) *Harmonic serialism*:

a. Given some input I_i, the candidate set CS_i = $\{O_{i1}, O_{i2}, ... O_{in}\}$ is generated by applying at most one operation to I_i.
b. The output O_{ij} with the best constraint profile is selected as optimal.
c. O_{ij} forms the input I_{ij} for the next generation step producing a new candidate set CS_j = $\{O_{ij1}, O_{ij2}, ... O_{ijn}\}$.
d. The output O_{ijk} with the best constraint profile is selected as optimal.
e. Candidate set generation stops (i.e., the derivation converges) when the output of an optimization procedure is identical to the input (i.e., when the constraint profile cannot be improved anymore).

In the present context, the main reason for adopting a harmonic serialist approach is that, in interaction with the PIC, it directly implements strict locality of constraint interaction: Since all competing outputs are separated from the input by at most one elementary operation, it can be ensured that there is no danger that processes taking place in potentially radically different areas of the sentence can interact with the process at issue in unwanted and unforeseen ways; in line with this, harmony evaluation based on weights assigned to constraints and to linguistic expressions remains feasible throughout since the number of interacting weights remains small.

3.4. Integrating Dependencies

Finally, it needs to be clarified how the optimization of structures involving extraction from NP can be made sensitive to $\Delta P_{V|N}$-based weight assignments to V-N dependencies. To this end, we postulate that X-Y dependencies relating to heads can function as syntactic primitives that constraints can refer to (and that they can restrict). This assumption has been made earlier in a number of otherwise quite different approaches, and sometimes with a different label attached to X-Y (like chains, catenae, or selections instead of dependency); see, e.g., O'Grady (1998), Osborne et al. (2013), Manzini (1995)), Bowers (2017), and Bruening (2018; 2020). For present purposes, we assume that dependencies (in this technical sense) are always two-membered (X-Y), and that they are characterized by a selection relation (X selects Y).[12] As detailed above, we assume that $\Delta P_{X|Y}$ determines the strength of an X-Y dependency. And we would like to suggest that the constraint where strength of dependencies plays a crucial role in the theory of extraction is the CONDITION ON EXTRACTION DOMAIN (CED; see Huang (1982), Chomsky (1986), and Cinque (1990)) in (25).

(25) CONDITION ON EXTRACTION DOMAIN (CED; violable, weighted):
For all X-Y dependencies, if X-Y intervenes between two adjacent members of a movement chain, X is a sister of the phrase headed by Y.

According to earlier versions of the CED, an XP blocks movement across it if it is not governed (see Huang (1982)), or not L(exically)-marked (see Chomsky (1986)), or not a complement (Cinque (1990)). It is this latter version that we adopt in (25). Furthermore, (25) formulates the CED as a constraint on X-Y dependencies intervening in a movement chain (rather than as a constraint on movement, or on adjacent members movement chains, as in the original versions). This is so as to ensure that it is the strength of the intervening X-Y dependency (rather than, say, the strength of the moved item, or of the movement chain that it is a part of) that determines CED satisfaction. Assuming the concept of intervention in (26), this change is innocuous.

(26) *Intervention*:[13]
An X-Y dependency intervenes between two members of a movement chain α_i and α_{i+1} iff (a) and (b) hold.

a. α_i m-commands X.
b. Y m-commands α_{i+1}.
c. It is not the case that X m-commands α_i and c-commands α_{i+1}.

Given (26), all but the most local instances of movement to either intermediate phase edges or final landing sites will cross an X-Y dependency.[14]

Based on these assumptions, we postulate that the CED plays a dual role in harmony evaluation. On the one hand, it is a negative constraint, just like MC and EC are: The

---

[12] For some cases (including, perhaps most notably, idiomatic expressions), it has been argued that dependencies can consist of more than two heads. Extending the present analysis from two-membered dependencies to n-membered dependencies would be possible; but this complication is not required for a proper treatment of extraction from NP, so we will not pursue the matter here.

[13] β m-commands γ if the next maximal projection dominating β also dominates γ (and β and γ are not in a dominance relation themselves); see Chomsky (1986). Note also that the third clause in (26) ensures that with movement to phase edge positions, the phase head itself cannot be part of an intervening dependency if it c-commands the lower member of a link of a movement chain (but it can be if the dependency goes into a specifier). This assumption is mainly made so as to reduce the number of intervening head-head dependencies, and keep evaluations as simple as possible. A version of the present approach where phase head dependencies always qualify as interveners would also be perfectly feasible.

[14] Extremely local movement steps are precluded in general by a designated Anti-Locality Condition; see Bošković (1997), Abels (2003), Grohmann (2003; 2011), Pesetsky (2016), and Erlewine (2016), among others. We will come back to this issue below.

CED registers a *violation* if it is violated by an output (and the strength of the violation depends on the strength of the X-Y dependency that gives rise to it). On the other hand, however, the CED is also a positive constraint, unlike MC and EC: It assigns a *reward* if it is satisfied. Positive constraints of this type are difficult to implement in standard parallel optimality theory (because of an Infinite Goodness problem arising according to which one could in principle carry out an infinite number of processes yielding rewards from a given constraint), but as noted by Kimper (2016), this problem vanishes under harmonic serialism, where input and output can be separated by at most one operation. Kimper observes that adopting positive constraint evaluation is empirically advantageous in the area of autosegmental spreading in phonology; and it turns out to also give rise to a much simpler account of the natural predicate effect with extraction from NP than would otherwise be available. Positive evaluation of the CED has the consequence that if an X-Y dependency satisfies the constraint, it can yield an additional reward, depending on the weight assigned to the X-Y dependency via $\Delta P_{X|Y}$.

3.5. Analyses

Let us look at some consequences. Suppose that MC is associated with a weight of 4.0, and EC with a weight of 5.0. Based on just these two constraints, the default consequence is that movement (or, in fact, any other kind of structure-building) is not possible: An output that carries out movement (in the presence of a designated feature [•F•]) will incur a violation (-1) of EC, and end up with a harmony value of -5.0. In contrast, a competing output that fails to apply movement will only trigger a violation (-1) of MC, therefore has an overall harmony value of -4.0 (other things being equal), and will thus always be selected. On this view, to bring about movement (i.e., to make the output with movement optimal), it is necessary to get a reward from the remaining constraint, CED.[15] We take the CED to be associated with a weight of 3.0.

Under these assumptions, a first prediction is that NP specifiers (subjects, indirect objects, and moved NPs) are invariably islands. Movement from a position within NP to the next edge of a phase will always violate the CED, and thus the bias against the movement-inducing MC will actually be strenghened. Note that there will be intervening dependencies in these environments – v selects the N head of a subject NP in Specv, V selects the N head of an indirect object NP in SpecV, and there can in principle also be a v-N dependency in the case of object scrambling to Specv.[16] Consequently, the CED springs into action here, and rules out extraction. This is shown for the case of extraction from a subject NP in (27).

[15] An analogous role must be played by some other constraint so as to make external Merge possible. There are several natural candidates for this, but we will remain silent on the issue in the present paper.

[16] As regards the first scenario of subject NPs in Specv, the v-N dependency intervenes for extraction from a subject NP in Specv to a higher Specv position because (a) α_i m-commands v, (b) N m-commands α_{i+1}, and (c) it is not the case that v both m-commands α_i and c-commands α_{i+1} – the former is the case but the latter is not; see (26). Essentially the same consequences arise for the third scenario of scrambled object NPs, given that intervention is determined without a recourse to base positions. In the second scenario involving indirect objects, there is only a V-N dependency that intervenes, as with direct objects.

(27) *Optimization of extraction from subject NP*:

I: $[_{vP}$ $[_{NP}$ ... $[_{N'}$ N ... XP_1]] $[_{VP}$ ... V] $v_{[\bullet X \bullet]}$]	MC w = 4.0	EC w = 5.0	CED w = 3.0	*H*
O_1: $[_{vP}$ XP_1 $[_{v'}$ $[_{NP}$ ... $[_{N'}$ N ... t_1]] $[_{VP}$ ... V] v]]		-1	-1	-8
☞O_2: $[_{vP}$ $[_{NP}$ ... $[_{N'}$ N ... XP_1]] $[_{VP}$ ... V] $v_{[\bullet X \bullet]}$]	-1			-4

In (27), output O_2 leaves XP_1 in situ, within the subject NP in Specv, even though, by assumption, there is a featural trigger for it. This gives rise to a -1 violation of the MC with weight 4.0, and to a harmony score of -4.0. On the other hand, O_1 extracts XP_1 out of the subject NP in Specv, to an outer Specv position, as required by MC (and ultimately by the PIC). This violates EC, yielding a violation score of -5.0. However, in addition, the CED is also violated since there is an intervening v-N dependency, and NP is not a sister of v. It is clear that, whatever the weight of the v-N dependency is, the constraint profile of the output that employs movement is thereby further worsened. For the sake of concreteness, we have registered a -1 violation of CED with O_1, yielding an overall harmony score of -8; but essentially the same result would have been obtained if the v-N dependency had a weight of, say, 0,01 (with -5,03 as the overall harmony score). The fact that the in-situ candidate O_2 wins this competition is, as such, not yet fatal. However, it is clear that XP_1 movement to the eventual target position later in the derivation will now eventually give rise to a fatal violation of the inviolable PIC.

Consider next the consequences that arise for extractions from NPs that are complements of V, i.e., direct objects. In this scenario, the CED is not violated. However, this does not yet suffice to permit extraction from the complement domain of N to the phase edge of v; in addition, there must be a sufficient reward from the CED (with weight 3.0) generated by an intervening V-N dependency. This reward may then render fatal the MC violation incurred by the output that does not apply movement, by lessening the EC violation incurred by the output that does. The reward is big enough in the well-formed cases of extraction from NP (i.e., where a natural predicate is involved), and too small in the ill-formed cases of NP (where V and N do not enter a tight relation).[17]

To illustrate this, we will focus on two weights assigned to V-N dependencies that are close to the dividing line between V-N dependencies that permit extraction and V-N depen-

[17] As a matter of fact, the idea that specific types of head-head dependencies can extend locality domains and permit extraction from XP which is otherwise blocked is not new. Koster (1987) postulates a Bounding Condition according to which each XP is a locality domain that as such blocks movement (and other processes), and that can only be made transparent by so-called "dynasties" of heads that stand in a government relation. Baker (1988) proposes that each XP is a priori a minimality barrier that can only be made transparent by movement of a head Y_1 to the next higher head X_2 that takes YP_1 as its sister; such head movement can be abstract (i.e., invisible), in which case it is signalled by a co-indexing of the two heads involved. Similarly, Staudacher (1990) suggests strengthening Chomsky's (1986) concept of L-marking to head-marking; on this view, a YP_1 is a barrier if it is not a complement of a head X_2 that specificically selects the head Y_1 of YP_1. Of course, none of these (and other, related) approaches can accomodate the frequency of V-N dependencies, in whatever form.

dencies that do not permit extraction; recall (19). Suppose first that the V-N dependency is equipped with a numerical value of 0.2. As shown in (28), this leads to a reward of 0.6 provided by the CED. Thus, the harmony score of the output that employs movement (i.e., O_1) is improved. However, (28) also shows that this does not yet suffice to license movement; the EC violation incurred by movement is still too strong, and leaving XP_1 in situ, as in O_2, remains the most harmonic strategy.

(28) *Optimization of extraction from direct object, $\Delta P_{V|N} \rightarrow 0.2$:*

I: $[_{vP}\ [_{VP}\ [_{NP} \ldots [_{N'}\ N \ldots XP_1\ ]]\ V\ ]\ v_{[\bullet X \bullet]}]$	MC w = 4.0	EC w = 5.0	CED w = 3.0	*H*
O_1: $[_{vP}\ XP_1\ [_{v'}\ [_{VP}\ [_{NP} \ldots [_{N'}\ N \ldots t_1\ ]]\ V\ ]\ v\ ]\ ]$		-1	+0.2	-4.4
☞O_2: $[_{vP}\ [_{VP}\ [_{NP} \ldots [_{N'}\ N \ldots XP_1\ ]]\ V\ ]\ v_{[\bullet X \bullet]}]$	-1			-4

Things are different when the V-N dependency has a weight of 0.4, though. As shown in (29), in this case the reduction effect brought about by the 1.2 reward for CED satisfaction is sufficiently large to permit the unavoidable violation of EC in the movement candidate O_1; and the MC violation incurred by the in-situ candidate O_2 becomes fatal.

(29) *Optimization of extraction from direct object, $\Delta P_{V|N} \rightarrow 0.4$:*

I: $[_{vP}\ [_{VP}\ [_{NP} \ldots [_{N'}\ N \ldots XP_1\ ]]\ V\ ]\ v_{[\bullet X \bullet]}]$	MC w = 4.0	EC w = 5.0	CED w = 3.0	*H*
☞O_1: $[_{vP}\ XP_1\ [_{v'}\ [_{VP}\ [_{NP} \ldots [_{N'}\ N \ldots t_1\ ]]\ V\ ]\ v\ ]\ ]$		-1	+0.4	-3.8
O_2: $[_{vP}\ [_{VP}\ [_{NP} \ldots [_{N'}\ N \ldots XP_1\ ]]\ V\ ]\ v_{[\bullet X \bullet]}]$	-1			-4

It is clear that all V-N dependencies with a weight higher than 0.4 (i.e., with higher $\Delta P_{V|N}$ values) will ceteris paribus also permit extraction from a complement NP, and that all V-N dependencies with a weight smaller than 0.2 will invariably block it. Thus, by assuming frequency-based ΔP values to act as weights associated with V-N dependencies, the concept of a natural predicate can be given a precise characterization, and asymmetries arising with extractions from NP in German can be derived.

3.6. Consequences

Needless to say, the present analysis makes a lot of further predictions, and raises several new questions. One obvious consequence is that not just extraction from NP, but in fact *all* instances of movement will depend on an intervening head-head dependency giving rise to a CED reward that sufficiently reduces the negative harmony value incurred by the EC violation inherent to movement, so as to make the output that carries out movement more harmonic than the output that does not (and that thereby violates MC). For instance, a movement step from Specv to SpecC (as in standard cases of wh-movement) crosses an intervening T-v dependency (though not a C-T dependency, given (26)). If nothing more is said, this dependency must be strong enough to bring about a sufficient CED reward to license the movement step, i.e., T-v must be associated with a weight of at least 0.34. We will assume that, more generally, when a head-head dependency involves two functional categories, or one functional category and one lexical category, the weight associated with it is typically very high; this follows naturally just by determining the ΔP values: A category like v is an extremely good predictor for a category like T, even if it is assumed that the particular phonological realizations of v and T are taken to be decisive (rather than the abstract functional category labels).

A further consequence of the analysis concerns EPP-driven movement of subject NPs to SpecT. Given a clause structure as in (30), there is no head-head dependency intervening between two members of a movement chain α_i and α_{i+1}.

(30) $[_{\text{TP}}\ \alpha_i\ [_{\text{T'}}\ \text{T}\ [_{\text{vP}}\ \alpha_{i+1}\ [_{\text{v}'}\ \text{v VP}\ ]]]]$

Consequently, the CED cannot be violated in (30), but there can also be no reward (since there is no dependency that satisfies the constraint non-trivially).[18] This means that movement should ceteris paribus be blocked in (30) (with the in-situ candidate violating MC being more harmonic than the movement candidate violating the constraint EC, which has a greater weight than MC). Several options suggest themselves to solve this problem. A simple solution would be that the EPP feature triggering (internal or external) Merge with T has more strength than other features triggering movement.[19]

Further consequences arise for extractions from NP that involve postposition stranding (see (1-b), (1-d)), i.e., extraction from PP in addition to extraction from NP. We have so far tacitly presupposed that N-P dependencies do not play a decisive role in licensing or prohibiting postposition stranding, but this might turn out to be the case upon closer scrutiny.

Finally, it can be noted that the present approach opens up the possibility of implementing Featherston's (2004) findings regarding the role of frequency in extraction from CPs in German in a very direct way. In German (and many other languages), the legitimacy of extraction from an embedded declarative clause headed by *dass* ('that') depends both on the grammatical function (direct object: yes, subject: no) and, more importantly in the present context, on the choice of matrix verb; only bridge verbs allow extraction. Two examples illustrating this are given in (31-a) (with bridge verbs) vs. (31-b) (with non-bridge verbs).

---

[18] In general, trivial constraint satisfaction by dependencies must not be able to generate a reward. Also note that, under the most natural concept of Anti-Locality (see footnote 14), the movement step in (30) would qualify as too local, necessitating additional assumptions to make it possible.

[19] In principle, assuming that instead of a single vP, there are actually two functional projections vP and VoiceP between TP and VP (see, e.g., Legate (2014) and Alexiadou et al. (2015)) would give rise to an intervening dependency that could in turn provide the additional CED reward licensing movement, and thus also solve the problem with EPP-driven movement. However, under this assumption, certain weight assignments would have to be changed in the core analysis above, given that there would then often be more than one intervening head-head dependency with movement to phase edge after all.

(31) a. (Ich weiß nicht) [$_{\text{CP}_1}$ wen$_4$ [$_{\text{vP}}$ t$_4'''$ sie meint/glaubt/sagt [$_{\text{CP}_2}$ t$_4''$ dass [$_{\text{vP}}$ t$_4'$
I know not whom she thinks/thinks/says that
du t$_4$ getroffen hast]]]]
you met have

b. ?*(Ich weiß nicht) [$_{\text{CP}_1}$ wen$_4$ [$_{\text{vP}}$ t$_4'''$ sie bereut/weiß/bezweifelt [$_{\text{CP}_2}$ t$_4''$ dass [$_{\text{vP}}$
I know not whom she regrets/knows/doubts that
t$_4'$ du t$_4$ getroffen hast]]]
you met have

Featherston's (2004) observation is that bridge verbs are more frequent than non-bridge verbs. Thus, the mean log frequencies of CP-embedding verbs that can be derived by collecting the absolute frequencies of these verbs in four different corpora, converting the numbers by applying a logarithm function, summing the four individual resulting numbers for each verb, and finally dividing them by four strongly correlate with the option of extraction from CP (which was determined by experiments involving grammaticality judgements). This is shown for the verbs *sagen* ('say'), *glauben* ('believe'), and *bezweifeln* ('regret') in (32).

(32)

Verb	Celex/spoken	Celex/written	Cosmas/spoken	Cosmas/written	Mean log freq.	Extract.
sagen	3545	8614	3637	93345	4.0039	-0.002
glauben	1076	1728	194	28877	3.2544	-0.155
bezweifeln	11	77	4	2241	1.7201	-0.764

Interestingly, even though Featherston (2004) has, in our view, convincingly identified frequency of the matrix verb as a factor determining the option of extraction from CP in German, the grammatical theory he employs (the Decathlon Model; see Featherston (2005; 2019)), while designed to predict frequencies in outputs, does in fact not incorporate frequency as a grammatical building block that may interact with other building blocks (like MC, EC, or CED in the present approach) to license or block extraction. Accordingly, Featherston (2004) remains silent on how to actually account for the frequency effect with the bridge verb phenomenon in grammatical theory. In contrast, it seems clear how the effect of frequency on extraction from CP could be modelled in the present approach. First, instead of bare V frequencies, $\Delta P_{V|C}$ values for V-C dependencies that intervene between a movement chain member α_{i+1} in SpecC and its immediate chain antecedent α_i in the matrix Specv have to be determined (we have not done this but are reasonably confident that the results will be very similar to Featherston's results). And second, scaled-up versions of these numbers are then predicted to bring about CED-based rewards that permit extraction from CP with highly frequent V-C dependencies (i.e., V-C dependencies that form a bridge).

4. Concluding Remarks

It is a standard observation that extraction of PPs and R-pronouns from direct object NPs in German is dependent on V and N forming a natural predicate. In this article, we have argued that this can and should be conceived of as a frequency effect: Only those V-N dependencies permit extraction from a direct object NP that have a sufficiently high $\Delta P_{V|N}$ value. In other words: Frequency can act as a language-external grammatical building block that transparently and directly interacts with language-internal grammatical building blocks regulating syntactic movement. We would like to contend that such a finding is difficult to reconcile with virtually all of the more widely adopted grammatical theories. It seems that the best one can do in standard approaches in order to implement the generalization is to view frequency as a factor determining the *learning* of syntactic operations, or rules. On such a view, highly frequent V-N dependencies could have become equipped with a special diacritic in the course of language acquisition, and the decision on whether movement can or cannot apply could then be made sensitive in the grammar to the presence or absence of this diacritic.[20] We take it to be uncontroversial that such a use of ad hoc diacritics whose sole purpose is to encode some other well-defined, independently existing piece of information that cannot be available in the grammar for systematic reasons is to be avoided if at all possible. As we have tried to show, Gradient Harmonic Grammar is unique among current theories of grammar in postulating that linguistic objects are associated with numerical weights that then interact with the weights assigned to the language-internal grammatical constraints, and that therefore make implementing frequency values a straightforward option. Our approach combines standard constraint evalulation of Gradient Harmonic Grammar with standard Minimalist derivations and standard Harmonic Serialism (which independently suggests itself for Minimalist derivations due to its inherently derivational nature). The only innovative assumption that we had to make is that the weights of V-N dependencies (as well as of other head-head dependencies) are determined by frequency.[21]

In addition to this substantive conceptual difference, a diacritic-based approach where frequency only plays a role in language acquisition and an approach where frequency acts as a language-external building block in the grammar itself are also not extensionally equivalent. At least in principle, they make different empirical predictions when it comes to *variation* in the domain of extraction from NP. Indeed, there seems to be quite a bit of variation with extraction from NP. In Gradient Harmonic Grammar, there are two natural sources for this: First, different *weights of constraints* (MC, EC, or CED, in the case at hand) can produce different optimal outputs. This implies that speakers with slightly different weights assigned to crucial constraints may simply have different thresholds for accepting or rejecting extraction from direct object NPs, without there being any weight differences with

[20] Arguably, the situation is basically identical in Construction Grammar, where entrenchment may make frequent V-N constructions amenable to extraction in the course of language acquisition; but frequency as such remains a factor relevant for *learning* a language here, and is not an actual *building block* active in the grammar (i.e., from a Construction Grammar perspective, the set of constructions exhibiting different degrees of abstractness, and the inheritance networks connecting them).

[21] In contrast, the weights of individual lexical items (and constituents more generally) in general do not seem to correspond to frequency; see Smolensky (2017), Lee (2018), Müller (2019), and also the earlier proposals in Ross (1973a;b; 1975).

respect to V-N dependencies. Second, different *weights of N-V dependencies* can of course also produce different optimal outputs. To end this article, it is this latter consequence that we would briefly like to focus on.[22]

Clearly, corpora like the DWDS core corpus can approximate the frequency of V-N dependencies in the external and internal linguistic inputs accessible to speakers. If the external linguistic input (i.e., the body of linguistic data outside of a speaker, which are accessible by hearing or reading) is vastly different, different outputs may become grammatical. To give a concrete example: Suppose that a speaker is immersed in a culture which is just like that of a prototypical German-speaking community, except that there is a tradition of of throwing books in the air after reading them. In that case, *Buch* ('book') will be a much better predictor for *werfen* ('throw') than it is in (18), and $\Delta P_{\text{werfen}|\text{Buch}}$ will be much higher. Here we may then expect that sentences like *Worüber hat Fritz ein Buch (in die Luft) geworfen?*, ('about what has Fritz a book (in the air) thrown') will become well formed. The same conclusion can be drawn for internal linguistic inputs (i.e., all the acts of thinking in terms of language without ever externalizing it, conducting inner monologues, and the like). Suppose, for instance, that some Nazi speaker fantasizes about burning books all the time and very clearly distinguishes between authors, or between topics, of the books that he wants to burn. In this scenario, $\Delta P_{\text{verbrennen}|\text{Buch}}$ will go up, and it would seem to be likely that this speaker will accept sentences like *Über wen soll ich heute ein Buch verbrennen?* ('about whom should I today a book burn'), which are certainly not well formed otherwise for most speakers (unless they have extremely reduced thresholds). These two thought experiments make it possible to distinguish empirically between the diacritic-based approach to frequency effects in extraction from NP and the purely frequency-based approach that we have pursued. In the former approach, frequency determines language acquisition and ceases to be active afterwards, whereas frequency stays active as a factor in the latter approach, and a change in frequency is expected to potentially lead to a change in the application of grammatical operations. Therefore, a change of the external linguistic input or of the internal linguistic input at any point in time is predicted to result in different extraction options under the direct approach to frequency effects advocated in the present paper, but not under the indirect approach that confines the role of frequency to language acquisition. Effects of the type hypothesized in this paragraph may then be taken as a further possible argument in support of the idea that frequency is directly active as a building block of grammar.

[22] A third possible source of variation arises if a stochastic component is added to the grammar; see, e.g., Hayes (2001), Bresnan et al. (2001), and Boersma & Pater (2016). We will not pursue this option here; the present system is strictly categorical.

References

Abels, Klaus (2003): Successive Cyclicity, Anti-Locality, and Adposition Stranding. PhD thesis, University of Connecticut, Storrs, Connecticut.

Alexiadou, Artemis, Elena Anagnastopoulou & Florian Schäfer (2015): *External Arguments in Transitivity Alternations*. Oxford University Press, Oxford.

Bach, Emmon & George Horn (1976): Remarks on 'Conditions on Transformations', *Linguistic Inquiry* pp. 265–299.

Baker, Mark (1988): *Incorporation. A Theory of Grammatical Function Changing*. University of Chicago Press, Chicago.

Baltin, Mark & Paul Postal (1996): More on Reanalysis Hypotheses, *Linguistic Inquiry* 27, 127–145.

Barbiers, Sjef (2002): Remnant Stranding and the Theory of Movement. In: *Dimensions of Movement*. Benjamins, Amsterdam, pp. 47–67.

Boersma, Paul & Joe Pater (2016): Convergence Properties of a Gradual Learning Algorithm for Harmonic Grammar. In: J. McCarthy & J. Pater, eds., *Harmonic Grammar and Harmonic Serialism*. Equinox, London, pp. 289–434.

Bošković, Željko (1997): *The Syntax of Nonfinite Complementation. An Economy Approach*. MIT Press, Cambridge, Mass.

Bowers, John (2017): *Deriving Syntactic Relations*. Cambridge University Press, Cambridge.

Bresnan, Joan, Shipra Dingare & Christopher Manning (2001): Soft Constraints Mirror Hard Constraints: Voice and Person in English and Lummi. In: *Proceedings of the LFG 2001 Conference*. CSLI Publications.

Broekhuis, Hans (2006): Derivations (MP) and Evaluations (OT). In: R. Vogel & H. Broekhuis, eds., *Optimality Theory and Minimalism: A Possible Convergence?*. Vol. 25, Linguistics in Potsdam, Potsdam, pp. 137–193.

Broekhuis, Hans & Ellen Woolford (2013): Minimalism and Optimality Theory. In: M. Den Dikken, ed., *Cambridge Handbook of Generative Syntax*. Cambridge University Press, Cambridge, pp. 122–161.

Bruening, Benjamin (2009): Selectional Asymmetries between CP and DP Suggest that the DP Hypothesis is Wrong. In: L. MacKenzie, ed., *Proceedings of the 32th Annual Penn Linguistics Colloquium*. Vol. 15.1 of *Penn Working Papers in Linguistics*, University of Pennsylvania, Philadelphia, pp. 26–35.

Bruening, Benjamin (2018): Idioms, Collocations, and Structure. Syntactic Constraints on Conventionalized Expressions. Ms., University of Delaware. To appear in *Natural Language and Linguistic Theory*.

Bruening, Benjamin (2020): The Head of the Nominal is N, not D: N-to-D Movement, Hybrid Agreement, and Conventionalized Expressions, *Glossa: A Journal of General Linguistics* 5(1), 15.

Bruening, Benjamin, Xuyen Dinh & Lan Kim (2018): Selection, Idioms, and the Structure of Nominal Phrases with and without Classifiers, *Glossa* 3, 1–46.

Cattell, Ray (1976): Constraints on Movement Rules, *Language* 52, 18–50.

Chomsky, Noam (1977): On Wh-Movement. In: P. Culicover, T. Wasow & A. Akmajian, eds., *Formal Syntax*. Academic Press, New York, pp. 71–132.

Chomsky, Noam (1986): *Barriers*. MIT Press, Cambridge, Mass.

Chomsky, Noam (1995): *The Minimalist Program*. MIT Press, Cambridge, Mass.

Chomsky, Noam (2001): Derivation by Phase. In: M. Kenstowicz, ed., *Ken Hale. A Life in Language*. MIT Press, Cambridge, Mass., pp. 1–52.

Chomsky, Noam (2008): On Phases. In: R. Freidin, C. Otero & M. L. Zubizarreta, eds., *Foundational Issues in Linguistic Theory*. MIT Press, Cambridge, Mass., pp. 133–166.

Cinque, Guglielmo (1990): *Types of A-bar Dependencies*. MIT Press, Cambridge, Mass.

Collins, Chris & Edward Stabler (2016): A Formalization of Minimalist Syntax, *Syntax* 19(1), 43–78.

Davies, William & Stanley Dubinsky (2003): On Extraction from NPs, *Natural Language and Linguistic Theory* 21, 1–37.

De Kuthy, Kordula (2001): Splitting PPs from NPs. In: W. D. Meurers & T. Kiss, eds., *Constraint-Based Approaches to Germanic Syntax*. CSLI Publications, Stanford University, pp. 25–70.

De Kuthy, Kordula & Walt Detmar Meurers (2001): On Partial Constituent Fronting in German, *Journal of Comparative Germanic Linguistics* 3(3), 143–205.

Diesing, Molly (1992): *Indefinites*. MIT Press, Cambridge, Mass.

Erlewine, Michael (2016): Anti-Locality and Optimality in Kaqchikel Agent Focus, *Natural Language and Linguistic Theory* 34, 429–479.

Fanselow, Gisbert (1987): *Konfigurationalität*. Narr, Tübingen.

Fanselow, Gisbert (1991): Minimale Syntax. Habilitation thesis, Universität Passau.

Fanselow, Gisbert (2001): Features, Theta-Roles, and Free Constituent Order, *Linguistic Inquiry* 32, 405–436.

Featherston, Sam (2004): Bridge Verbs and V2 verbs: The Same Thing in Spades?, *Zeitschrift für Sprachwissenschaft* 23, 181–210.

Featherston, Sam (2005): The Decathlon Model of Empirical Syntax. In: S. Kepser & M. Reis, eds., *Linguistic Evidence*. Mouton de Gruyter, Berlin, pp. 187–208.

Featherston, Sam (2019): The Decathlon Model. In: A. Kertész, C. Rákosi & E. Moravcsik, eds., *Current Approaches to Syntax – a Comparative Handbook*. Mouton de Gruyter, Berlin, pp. 155–186.

Frey, Werner (2015): NP-Incorporation in German. In: O. Borik & B. Gehrke, eds., *The Syntax and Semantics of Pseudo-Incorporation*. Brill, Leiden, pp. 225–261.

Gallego, Ángel (2007): Phase Theory and Parametric Variation. PhD thesis, Universitat Autónoma de Barcelona, Barcelona.

Gazdar, Gerald & Ewan Klein (1978): Review of Ed Keenan, ed., Formal Semantics of Natural Language, *Language* 54, 661–667.

Georgi, Doreen (2017): Patterns of Movement Reflexes as the Result of the Order of Merge and Agree, *Linguistic Inquiry* 48(4), 585–626.

Georgi, Doreen & Gereon Müller (2010): Noun Phrase Structure by Reprojection, *Syntax* 13(1), 1–36.

Geyken, Alexander (2007): The DWDS Corpus: A Reference Corpus for the German Language of the 20th Century. In: C. Fellbaum, ed., *Collocations and Idioms: Linguistic, Lexicographic, and Computational Aspects*. Continuum Press, London, pp. 23–41.

Grewendorf, Günther (1989): *Ergativity in German*. Foris, Dordrecht.

Gries, Stefan (2013): 50-Something Years of Work on Collocations: What Is or Should Be Next, *International Journal of Corpus Linguistics* 18(1).

Gries, Stefan, Beate Hampe & Doris Schönefeld (2005): Converging Evidence: Bringing Together Experimental and Corpus Data on the Association of Verbs and Constructions, *Cognitive Linguistics* 16, 635–676.

Gries, Stephan & Anatol Stefanowitsch (2004): Extending Collostructional Analysis. A Corpus-Based Perspective on "Alternations", *International Journal of Corpus Linguistics* 9, 97–129.

Grimshaw, Jane (1997): Projection, Heads, and Optimality, *Linguistic Inquiry* 28, 373–422.

Grimshaw, Jane (2006): Chains as Unfaithful Optima. Ms., Rutgers University. Also in *Wondering at the Natural Fecundity of Things: Essays in Honor of Alan Prince*; ROA 844, 97-110.

Grohmann, Kleanthes K. (2003): Successive Cyclicity under (Anti-)Local Considerations, *Syntax* 6, 260–312.

Grohmann, Kleanthes K. (2011): Anti-Locality: Too-Close Relations in Grammar. In: C. Boeckx, ed., *The Oxford Handbook of Linguistic Minimalism*. Oxford University Press, pp. 260–290.

Haider, Hubert (1983): Connectedness Effects in German, *Groninger Arbeiten zur Germanistischen Linguistik* 23, 82–119.

Haider, Hubert (1993): *Deutsche Syntax – generativ*. Narr, Tübingen.

Halle, Morris & Alec Marantz (1993): Distributed Morphology and the Pieces of Inflection. In: K. Hale & S. J. Keyser, eds., *The View from Building 20*. MIT Press, Cambridge, Mass., pp. 111–176.

Hayes, Bruce (2001): Gradient Well-Formedness in Optimality Theory. In: J. Dekkers, F. van der Leeuw & J. van de Weijer, eds., *Optimality Theory. Phonology, Syntax, and Acquisition*. Oxford University Press, Oxford, pp. 88–120.

Heck, Fabian & Gereon Müller (2013): Extremely Local Optimization. In: H. Broekhuis & R. Vogel, eds., *Linguistic Derivations and Filtering*. Equinox, Sheffield, pp. 135–166.

Hein, Johannes & Katja Barnickel (2018): Replication of R-pronouns in German Dialects, *Zeitschrift für Sprachwissenschaft* 37, 171–204.

Huang, Cheng-Teh James (1982): Logical Relations in Chinese and the Theory of Grammar. PhD thesis, MIT, Cambridge, Mass.

Kimper, Wendell (2016): Positive Constraints and Finite Goodness in Harmonic Serialism. In: J. McCarthy & J. Pater, eds., *Harmonic Grammar and Harmonic Serialism*. Equinox, London, pp. 221–235.

Koster, Jan (1987): *Domains and Dynasties*. Foris, Dordrecht.

Lee, Hyunjung (2018): Generalized Complementizer-Trace Effects in Gradient Harmonic Grammar: Deriving Extraction Asymmetries. Ms., Universität Leipzig. (Talk, 40th DGfS Meeting, Stuttgart, March 2018).

Legate, Julie Anne (2014): *Voice and v. Lessons from Acehnese.* MIT Press, Cambridge, Mass.

Legendre, Géraldine, Colin Wilson, Paul Smolensky, Kristin Homer & William Raymond (2006): Optimality in Syntax II: Wh-Questions. In: P. Smolensky & G. Legendre, eds., *The Harmonic Mind.* Vol. II, MIT Press, Cambridge, Mass., chapter 14, pp. 183–230.

Mahajan, Anoop (1992): The Specificity Condition and the CED, *Linguistic Inquiry* 23, 510–516.

Manzini, Rita (1995): From Merge and Move to Form Dependency, *UCL Working Papers in Linguistics* pp. 323–345.

Marantz, Alec (1995): 'Cat' as a Phrasal Idiom: Consequences of Late Insertion in Distributed Morphology. Ms., MIT, Cambridge, Mass.

McCarthy, John (2010): An Introduction to Harmonic Serialism, *Language and Linguistics Compass* 4, 1001–1018.

McCarthy, John & Joe Pater, eds. (2016): *Harmonic Grammar and Harmonic Serialism.* Equinox, Sheffield.

Müller, Gereon (1991): Abstrakte Inkorporation. In: S. Olsen & G. Fanselow, eds., *DET, COMP und INFL.* Niemeyer, Tübingen, pp. 155–202.

Müller, Gereon (1995): *A-bar Syntax.* Mouton/de Gruyter, Berlin.

Müller, Gereon (1998): *Incomplete Category Fronting.* Kluwer, Dordrecht.

Müller, Gereon (2011): *Constraints on Displacement. A Phase-Based Approach.* Vol. 7 of *Language Faculty and Beyond*, Benjamins, Amsterdam.

Müller, Gereon (2019): The Third Construction and Strength of C: A Gradient Harmonic Grammar Approach. In: K. R. Christensen, H. Jørgensen & J. L. Wood, eds., *The Sign of the V – Papers in Honour of Sten Vikner.* Aarhus University: Dept. of English, School of Communication and Culture, Aarhus, pp. 419–448.

Murphy, Andrew (2017): Cumulativity in Syntactic Derivations. PhD thesis, Universität Leipzig.

Newmeyer, Frederick (1986): *Linguistic Theory in America.* The University of Chicago Press, Chicago.

Newmeyer, Frederick (2000): *Language Form and Language Function.* MIT Press, Cambridge, Mass.

O'Grady, William (1998): The Syntax of Idioms, *Natural Language and Linguistic Theory* 16, 279–312.

Osborne, Timothy, Michael Putnam & Thomas Groß (2013): Catenae: Introducing a Novel Unit of Syntactic Analysis, *Syntax* .

Ott, Dennis (2011): A Note on Free Relative Clauses in the Theory of Phases, *Linguistic Inquiry* 42, 183–192.

Pater, Joe (2009): Weighted Constraints in Generative Linguistics, *Cognitive Science* 33, 999–1035.

Pater, Joe (2016): Universal Grammar with Weighted Constraints. In: J. McCarthy & J. Pater, eds., *Harmonic Grammar and Harmonic Serialism*. Equinox, Sheffield, pp. 1–46.

Pesetsky, David (2016): Exfoliation: Towards a Derivational Theory of Clause Size. Ms., MIT, Cambridge, Mass.

Pesetsky, David & Esther Torrego (2006): Probes, Goals and Syntactic Categories. Ms., MIT. Lingbuzz/000321.

Prince, Alan & Paul Smolensky (1993): Optimality Theory. Constraint Interaction in Generative Grammar. Book ms., Rutgers University.

Prince, Alan & Paul Smolensky (2004): *Optimality Theory. Constraint Interaction in Generative Grammar*. Blackwell, Oxford.

Putnam, Michael & Lara Schwarz (2017): Predicting the Well-Formedness of Hybrid Representations in Emergent Production. Ms., Penn State University.

Riemsdijk, Henk van (1978): *A Case Study in Syntactic Markedness: The Binding Nature of Prepositional Phrases*. Foris, Dordrecht.

Ross, John (1973a): A Fake NP Squish. In: C.-J. Bailey & R. Shuy, eds., *New Ways of Analyzing Variation in English*. Georgetown University Press, Washington, DC, pp. 96–140.

Ross, John (1973b): Nouniness. In: O. Fujimura, ed., *Three Dimensions of Linguistic Research*. TEC Company Ltd, Tokyo, pp. 137–257. Reprinted in Bas Aarts et al. (eds.): *Fuzzy Grammar: A Reader*, 351-422. Oxford: Oxford University Press.

Ross, John (1975): Clausematiness. In: E. Keenan, ed., *Formal Semantics of Natural Language*. Cambridge University Press, New York and Cambridge, pp. 422–475.

Sauerland, Uli (1995): Review of 'A-bar Syntax' by Gereon Müller, *Glot International* 1(8), 14–15.

Schmellentin, Claudia (2006): *PP-Extraktionen. Eine Untersuchung zum Verhältnis von Grammatik und Pragmatik*. Niemeyer, Tübingen.

Schwarz, Lara (2020): Accepting Our Mistakes: How Variation Completes the Linguistic Puzzle. Talk, DGfS 42, Hamburg.

Smolensky, Paul (2017): Gradient Representations. Tutorial, Universität Leipzig; November 10-12, 2017.

Smolensky, Paul & Geraldine Legendre (2006): *The Harmonic Mind*. MIT Press, Cambridge, Mass.

Smolensky, Paul & Matthew Goldrick (2016): Gradient Symbolic Representations in Grammar: The Case of French Liaison. ROA 1286.

Staudacher, Peter (1990): Long Movement from Verb-Second-Complements in German. In: G. Grewendorf & W. Sternefeld, eds., *Scrambling and Barriers*. Benjamins, Amsterdam, pp. 319–339.

Stefanowitsch, Anatol (2009): Bedeutung und Gebrauch in der Konstruktionsgrammatik. Wie kompositionell sind modale Infinitive im Deutschen?, *Zeitschrift für Germanistische Linguistik* 37, 565–592.

Sternefeld, Wolfgang (1991): *Syntaktische Grenzen. Chomskys Barrierentheorie und ihre Weiterentwicklungen*. Westdeutscher Verlag, Opladen.

Trissler, Susanne (1993): P-Stranding im Deutschen. In: F.-J. e. a. d'Avis, ed., *Extraktion im Deutschen I*. Vol. 34, Arbeitspapiere des SFB 340, Stuttgart & Tübingen, pp. 247–291.

Urk, Coppe van (2015): A Uniform Syntax for Phrasal Movement: A Dinka Bor Case Study. PhD thesis, MIT, Cambridge, Mass.

Webelhuth, Gert (1988): A Universal Theory of Scrambling. In: *Papers from the 10th Scandinavian Conference on Linguistics*. Vol. ii, Dept. of Linguistics and Phonetics, University of Bergen, pp. 284–298.

Webelhuth, Gert (1992): *Principles and Parameters of Syntactic Saturation*. Oxford University Press, Oxford.