Gradient HPSG

Berke Şenşekerci
University of Warsaw

Proceedings of the 31st International Conference on
Head-Driven Phrase Structure Grammar

Palacký University Olomouc, Czech Republic

Stefan Müller, Rui Chaves (Editors)

2024

Frankfurt/Main: University Library

pages 173–192

Keywords: HPSG, gradience, coordination of unlikes, experimental syntax

Şenşekerci, Berke. 2024. Gradient HPSG. In Stefan Müller & Rui Chaves (eds.), *Proceedings of the 31st International Conference on Head-Driven Phrase Structure Grammar, Palacký University Olomouc, Czech Republic*, 173–192. Frankfurt/Main: University Library. DOI: 10.21248/hpsg.2024.11.

Abstract

Prevailing grammatical frameworks treat grammaticality as a binary concept, despite strong experimental evidence suggesting it is better understood as a gradient notion. This highlights a serious disconnect between linguistic theory and empirical data. While a few truly gradient frameworks have been proposed to bridge the gap, none have been developed within a constraint-based formalism – an approach particularly well-suited for modeling gradient grammaticality. This work formally introduces a gradient version of HPSG and subsequently employs it to analyze acceptability judgment data on unlike coordination phenomena in Turkish, which display distinctly gradient patterns.

1 Introduction

The notion that grammaticality of sentences cannot be neatly divided into two categories has been recognized since the early days of generative linguistics (Bolinger 1961, Chomsky 1961, Chomsky 1965: 148–153) and has found consistent support in subsequent work involving controlled acceptability judgment experiments (Keller 2000, Keller & Alexopoulou 2001, Featherston 2005b, Sorace & Keller 2005, Haegeman et al. 2014, Hofmeister et al. 2014).[1]

Despite substantial support in favor of gradience, prevailing grammatical frameworks persist in upholding a binary view of grammaticality, forcing linguists to rely on arbitrary generalizations when interpreting acceptability judgment data. As a result, these frameworks allow vastly different grammars to emerge from the same data depending on the chosen cutoff point between grammatical and ungrammatical.

To remedy this problem, various proposals have been put forward. Notable among these are Harmonic Grammar (Legendre et al. 1990), Linear Optimality Theory (Keller 2000), and the Decathlon Model (Featherston 2005a).[2] Interestingly, no such attempt has been made within a fully-fledged

[†]I want to thank Adam Przepiórkowski, whose guidance was invaluable during the conception of this work, and Marcin Opacki for reviewing an earlier version. I also extend my thanks to the audience at HPSG 2024 and to the anonymous reviewers. Any remaining errors are my own.

[1]I tentatively attribute the gradience observed in controlled acceptability judgment experiments to gradience in grammar and not to processing effects, as these experiments significantly minimize and control for such confounds. Accordingly, throughout this paper, I refer to gradient *grammaticality*, not gradient *acceptability*. Nevertheless, it is worth noting that the distinction between performance and grammar effects on acceptability has not been thoroughly explored in experimental settings (but see Hofmeister et al. 2014).

[2]Although Optimality Theory and stochastic variants of existing frameworks may appear to be viable options for modeling gradience, they have notable shortcomings. Optimality Theory not only presupposes binary grammaticality but is also fundamentally incompatible with judgment data (Keller & Asudeh 2002). Similarly, stochastic frameworks are specifically designed to model corpus frequencies, which are a distinct type of data that should not be conflated with acceptability judgments (Pullum & Scholz 2001: 31).

constraint-based framework like Head-Driven Phrase Structure Grammar (Pollard & Sag 1994, Müller et al. 2024), although a constraint-based backbone has been considered to be especially suitable for modeling gradience (Pullum & Scholz 2001: §3.1, Sag & Wasow 2011, Wasow 2024).

In light of this gap, the present work proposes a version of HPSG that accommodates the gradient grammaticality observed in acceptability judgment experiments. Subsequently, the proposed framework is utilized to analyze the results of an acceptability judgment experiment investigating unlike coordination phenomena in Turkish.

2 Gradient HPSG

Compelling evidence suggests that the grammaticality of a sentence is a matter of degree, primarily determined by two distinct factors (Keller 2000, Featherston 2005a, Sorace & Keller 2005): 1) the number of violations present, and 2) the relative severity of the violated constraints. To model gradient grammaticality in terms of these two factors, Gradient HPSG introduces two modifications to the model theory of HPSG.[3]

The first modification updates the original definition of an HPSG grammar (Richter 2004: 178) to allow each grammar constraint to be assigned a numeric weight that reflects the severity of its violation:

Definition 1 (grammar) Γ *is a grammar iff*
Γ *is a pair* $\langle \Sigma, \theta \rangle$,
Σ *is a septuple* $\langle S, \sqsubseteq, S_{max}, A, F, R, Ar \rangle$,
θ *is a set of ordered pairs such that:*
$\theta = \{\langle \delta, w \rangle \mid \delta \in D_0^{\Sigma} \wedge w \in \mathbb{R}^+\}$

The original definition of a *signature*, denoted by Σ (Richter 2004: 156), remains unchanged. This essentially means that Gradient HPSG does not introduce gradience to type hierarchies (cf. Brew 1995). However, θ is no longer a set of constraints as originally defined, but instead a set of ordered pairs where each pair consists of a constraint, δ, and its weight, w, which can only be a positive real number.

The second modification concerns the definition of a *model*, which originally classifies a sentence as a well-formed structure within a *model* of a grammar *iff* the sentence satisfies each constraint of the grammar (Richter 2004: 178–179).

By contrast, Gradient HPSG posits that the modelness (or well-formedness) of a sentence is a real number from 0 to negative infinity, where sentences

[3] Throughout the paper, 'model theory of HPSG' refers to *Relational Speciate Reentrant Logic* (RSRL; Richter 2004).

with 0 modelness value are perfect models of the grammar – i.e., they do not violate any constraint of the grammar.

As per the two factors underlying the grammaticality of a sentence, this value is determined on the basis of constraint weights and the number of constraint violations present in a sentence. The following definition of a *model* assumed in Gradient HPSG formalizes this concept (to be revised):

Definition 2 (model; preliminary) *For each grammar* $\Gamma = \langle \Sigma, \theta \rangle$ *and for each* Σ *interpretation* $I = \langle U, S, A, R \rangle$
Modelness degree of I *with respect to* Γ *is:*

$$M(I) = -\sum_{\langle \delta_i, w_i \rangle \in \theta} |U \backslash D_I(\delta_i)| \cdot w_i$$

The mathematical function that determines the modelness degree of a sentence is conceptually equivalent to the harmony function operationalized in Linear Optimality Theory (Keller 2000: 253): it computes the weighted sum of constraint violations for each constraint δ_i in a grammar. However, the function used in this definition is model-theoretic, operating strictly on HPSG structures.

The first term following the negated summation, $|U \backslash D_I(\delta_i)|$, returns the number of entities that are *not* denoted by δ_i. In simpler terms, this term counts the number of violations that a sentence makes with respect to δ_i. The number of δ_i violations obtained by this term is subsequently multiplied by the weight assigned to δ_i, w_i. For example, if a sentence violates δ_i twice and the weight of δ_i is specified as 0.45 in the grammar, the sentence receives an evaluation of 0.90 with respect to δ_i (2×0.45).

This evaluation procedure is carried out for each and every constraint in the grammar, with the outcomes of each assessment summed. The resulting sum is then negated to render the modelness value more intuitive, as higher values obtained from the weighted sum indicate greater degrees of ill-formedness rather than well-formedness.

Alas, this definition does not work as intended on standard RSRL assumptions regarding the shape of models: it does not compute the modelness degree of an individual sentence with respect to the weighted grammar. The standard model theory of HPSG posits that models reflect language as a whole (King 1999, Richter 2004, 2007), i.e., that they are *exhaustive models*. This assumption implies that models include all possible sentences within a language, as well as various partial HPSG objects, such as SYNSEM objects. Consequently, the function presented in Definition 2 iterates over all such objects instead of a specific sentence.

To ensure that this function takes an individual sentence as its input, Gradient HPSG additionally incorporates Przepiórkowski's (2021) revisions to the model-theory of HPSG that restrict models to correspond strictly to individual sentences (i.e., rooted, non-exhaustive models).[4] The final definition is presented below, with *interpretation* I now formally defined as a 5-tuple that includes a root element, denoted as r:

Definition 3 (model) *For each grammar* $\Gamma = \langle \Sigma, \theta \rangle$ *and for each* Σ *interpretation* $I = \langle U, r, S, A, R \rangle$
Modelness degree of I *with respect to* Γ *is:*

$$M(I) = -\sum_{\langle \delta_i, w_i \rangle \in \theta} |U \backslash D_I(\delta_i)| \cdot w_i$$

Having established the formal properties of Gradient HPSG, we can now proceed to illustrate its application in the formal analysis of acceptability judgment data.

3 Experiment

3.1 Background

The morphosyntactic properties of coordinate structures have been the subject of prolonged debate. One widely adopted position contends that conjuncts must bear the same syntactic category (Chomsky 1957: 36, Williams 1981: §2, Bruening & Al Khalaf 2020) and grammatical case (Weisser 2020).

Counter-examples to this position, where conjuncts mismatch either in their category, such as (1a)–(1b), or case, as in (1c), have been explained away by invoking various analytical mechanisms, such as supercategories (Bruening & Al Khalaf 2020), ellipsis (Beavers & Sag 2004: 54–56), and allomorphy (Weisser 2020: §2.3).

(1) a. Pat is [[$_{\text{NP}}$ a Republican] and [$_{\text{ADJP}}$ proud of it]].
(Sag et al. 1985: 117, ex. (2b))

b. We walked [[$_{\text{ADVP}}$ slowly] and [$_{\text{PP}}$ with great care]].
(Sag et al. 1985: 140, ex. (57))

c. This is starting to make [him and I] both feel really bad.
(Parrot 2009: 274, ex. (7a))

This position has recently been challenged based on an abundance of attested examples from Polish and English that defy such analyses (Patejuk 2015, Dalrymple 2017, Przepiórkowski 2022, Patejuk & Przepiórkowski 2023),

[4] The analysis in Section 4 further adopts the second-order extension of HPSG's model theory proposed by Przepiórkowski (2021).

suggesting a potential collapse of this generalization in the face of cross-linguistic evidence.

The current debate, however, is limited to English and Polish data. To further challenge this position through an experimental paradigm, a formal acceptability judgment experiment was conducted to gather data from Turkish, an agglutinative and head-final language.[5]

3.2 Methodology

In the experiment, 48 native speakers of Turkish evaluated the acceptability of sentences on a 7-point Likert scale from -3 (completely unnatural) to 3 (completely natural).[6] The experimental hypothesis posited that conjoining unlike categories and cases is acceptable in Turkish, provided that the conjuncts share the same grammatical function.

The experimental design consisted of two blocks: one for unlike categories and another for unlike cases. The category block had a standard 2×2 design, where the two crossed factors were the category of conjuncts (like or unlike: LCAT vs. UCAT) and the grammatical function of conjuncts (like or unlike: LF vs. UF). For the case block, a similar design was pursued – like or unlike cases (LCASE vs. UCASE) and grammatical functions (LF vs. UF). However, in this block, only three levels were feasible, as the construction of LCASE-UF stimuli was limited by the strict mapping between cases and grammatical functions in Turkish.

Sentence stimuli were constructed using the token-set methodology (Cowart 1997). This resulted in 12 token sets per block and a total of 84 sentences ($12 \times 4 + 12 \times 3$). All stimuli were based on examples of unlike coordination extracted from the Turkish Web 2012 corpus (Baisa & Suchomel 2012). To minimize attrition effects, the materials were split into 4 sub-surveys following the Latin square method. As a result, each participant saw 21 target sentences, along with 22 uncontroversially grammatical or ungrammatical fillers and 3 practice sentences.

3.3 Results

3.3.1 Category block

In the 12 token sets in the category block, the UCAT-LF sentences crucial for the hypothesis contained different categories of adjuncts (9 sentences with different categories selected from: AdvP, NP, and PP), arguments (2

[5]I would like to acknowledge the assistance I received from Adam Przepiórkowski, Katarzyna Kuś, Erkan Şenşekerci, and Szymon Talaga during the implementation of the experiment.

[6]This experiment is also described in Şenşekerci & Przepiórkowski (2024), which proposes an LFG analysis of the relevant data under the assumption of binary grammaticality.

sentences of "PP & NP" coordinations), and predicates (1 sentence of "NP & AP" coordination).

As shown in Figure 1, such UCAT-LF sentences received high scores on average. While LCAT-LF sentences, which featured fully parallel coordinations, were rated slightly higher than UCAT-LF sentences, this difference did not reach statistical significance ($p = .11$). A sharp decline in acceptability was observed only in LCAT-UF and UCAT-UF sentences ($p < .001$ w.r.t. UCAT-LF), where the conjuncts had different grammatical functions.

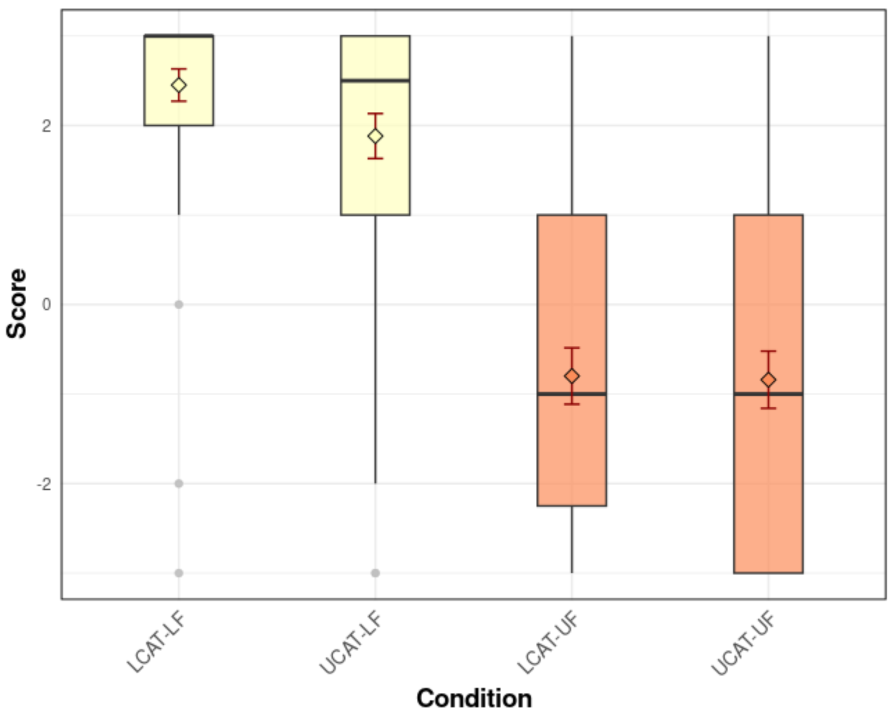

Figure 1: Raw scores of the category block stimuli (y-axis) by sentence type (x-axis), with means indicated by diamonds, and 95% confidence intervals of means by red error bars.

3.3.2 Case block

In the case block, the 12 UCASE-LF sentences with unlike cases but identical adjunct grammatical functions each incorporated cases typical for NP adjuncts: ablative, instrumental, and locative. For example, 4 sentences had coordinations of the type "NP-LOC & NP-ABL".

As shown in Figure 2, these UCASE-LF sentences received significantly lower, yet still positive, judgments compared to LCASE-LF sentences ($p < .001$).

Consistent with the category block results, the average acceptability dropped below zero only for UCASE-UF sentences ($p < .001$ w.r.t. UCASE-LF).

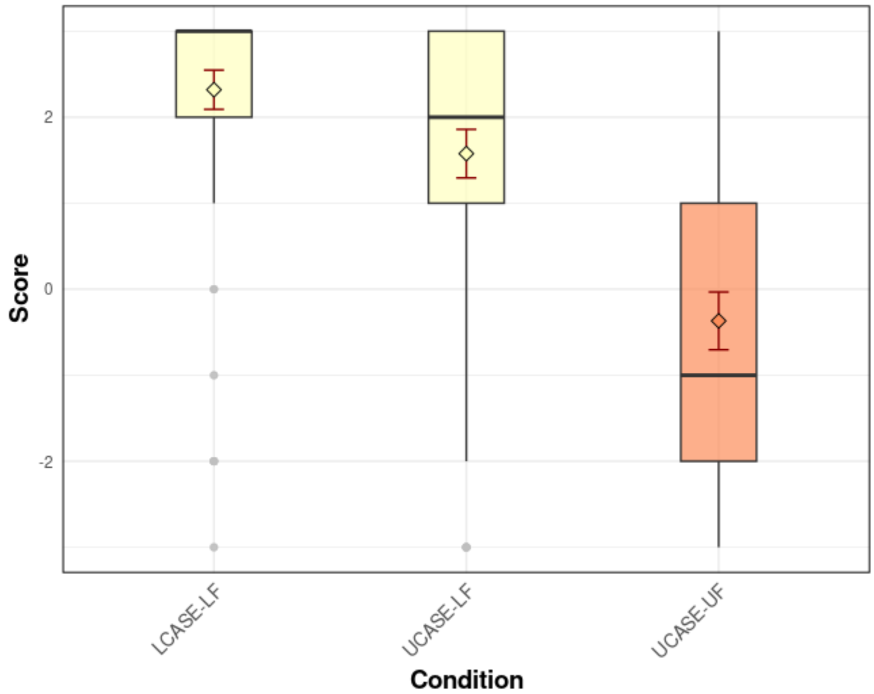

Figure 2: Raw scores of the case block stimuli (y-axis) by sentence type (x-axis), with means indicated by diamonds, and 95% confidence intervals of means by red error bars.

In summary, the results from both experimental blocks support the hypothesis: UCAT-LF and UCASE-LF types of coordination are acceptable. Nevertheless, the fact that such types are not as acceptable as their fully parallel counterparts (i.e., LCAT-LF and LCASE-LF) necessitates a gradient analysis to fully account for the empirical observations.

4 Analysis

As pointed out in the previous section, both coordination of unlike arguments and adjuncts were tested in UCAT-LF and UCASE-LF sentences. Both configurations (i.e., unlike arguments and unlike adjuncts) are acceptable due to the very same reason: satisfaction of disjunctive selectional requirements. Unlike arguments meet the disjunctive requirements imposed on them, while unlike adjuncts modify heads that satisfy the requirements of adjuncts themselves. However, the formal constraints that account for them are different.

4.1 Coordination of unlike arguments

In the case of coordination of unlike arguments, the relevant generalization pertains to the disjunctive requirements imposed by the predicate on the HEAD values of its complements.

For instance, the predicate *sür-* 'last/continue' takes a nominative NP as its subject and a durative complement that can be 1) a nominative NP, as in (2a); 2) a PP projected either by the postposition *boyunca* 'throughout', as in (2b), or *kadar* 'until', as in (2c); or 3) an AdvP, as in (2d).

(2) a. Tahliye çalışma-lar-ı *iki saat* sür-dü.
evacuation work-PL-3.POSS two hour.NOM last-PST
'The evacuation efforts lasted two hours.'

b. Bu kısır döngü *ilk 45 dakika boyunca* sür-dü.
this infertile cycle first 45 minute throughout last-PST
'This vicious cycle continued for the first 45 minutes.'

c. Bu süreç *nisan ay-ı-na kadar* sür-dü.
this phase april month-3P-DAT until last-PST
'This phase lasted until April.'

d. Onlar-ın etki-si *yıl-lar-ca* sür-er.
they-GEN effect-3P year-PL-ADVZ last-AOR
'Their effect lasts for years.'
(Turkish Web 2012; Baisa & Suchomel 2012)

While the coordinated subjects of this verb must be strictly parallel (i.e., all must be nominative NPs), the coordinated complements may mismatch as long as each coordinand satisfies one of the requirements imposed by *sür-*, as in (3).

(3) Bu program [[$_{\text{NP}}$ her hafta] ve [$_{\text{ADVP}}$ saat-ler-ce]] sür-ecek.
this program every week and hour-PL-ADVZ last-FUT
'This program will run every week and for hours.'
(Turkish Web 2012)

To ensure that these selectional requirements are evaluated individually for each conjunct, we employ the c relation (Yatabe 2004, Przepiórkowski 2021), defined in (4). This relation accepts an object and a description as input and checks whether the description holds true for the object.[7] If the

[7] Note that relations can accept descriptions as their inputs in second-order HPSG (Przepiórkowski 2021: 174–178).

input object is a coordination, the relation checks the description against each element in the ARGS list, which contains the HEAD values of each conjunct.

(4) $\forall \boxed{1}_e\ \forall \alpha_{et}\ (\ \mathsf{c}(\boxed{1}, \alpha) \leftrightarrow \alpha(\boxed{1})\ \ \vee$
$\exists \boxed{a_1} \ldots \exists \boxed{a_n}\ (\ \boxed{1}\big[\text{ARGS}\ \big\langle \boxed{a_1}, \ldots, \boxed{a_n} \big\rangle\big]\ \wedge$
$\mathsf{c}(\boxed{a_1}, \alpha) \wedge \ldots \wedge \mathsf{c}(\boxed{a_n}, \alpha)\ )\ )$

(Przepiórkowski 2021: 177, ex. (21))

Accordingly, a (simplified) lexical entry for *sür-* can be formalized as shown in (5) where the selectional requirements of *sür-* are checked separately for subject and object position via c relation.[8]

(5)
$$\begin{bmatrix} \text{PHON} & \langle \textit{sür} \rangle \\ \text{SYNSEM}|\text{CAT}|\text{VALENCE} & \begin{bmatrix} \text{SUBJ} & \big\langle \big[\text{CAT}|\text{HEAD}\ \boxed{1}\big] \big\rangle \\ \text{COMPS} & \big\langle \big[\text{CAT}|\text{HEAD}\ \boxed{2}\big] \big\rangle \end{bmatrix} \end{bmatrix}$$
$$\wedge\ \alpha_1 \approx (:\sim \textit{noun} \wedge :\text{CASE} \sim \textit{nom})$$
$$\wedge\ \alpha_2 \approx [(:\sim \textit{noun} \wedge :\text{CASE} \sim \textit{nom}) \vee (:\sim \textit{postp} \wedge (:\text{PFORM} \sim \textit{boyunca} \vee :\text{PFORM} \sim \textit{kadar})) \vee (:\sim \textit{adv})]$$
$$\wedge\ \mathsf{c}(\boxed{1}, \alpha_1) \wedge \mathsf{c}(\boxed{2}, \alpha_2)$$

4.2 Coordination of unlike adjuncts

An analogous analysis can be applied to unlike adjuncts. However, under standard HPSG assumptions, modifiers select for their heads, which necessitates encoding such disjunctive requirements within the lexical entries of modifiers.

Experimental findings and a related corpus investigation indicate that verbal heads can be modified by 1) any PP, as in (6a); 2) any AdvP, as in (6b); 3) NPs in locative, ablative, or instrumental case, as in (6c); or 4) a coordination of these options, which may involve unlike coordination as in the attested (6d).

[8]The constraint in (5) employs two RSRL operators: '∼' and ':', which are sort assignment and identity functions, respectively. For example, the RSRL description ':CASE ∼ *nom*' denotes those objects where the given path, CASE, leads to an object of sort *nom*. Accordingly, the description assigned to α_1 – $(:\sim \textit{noun} \wedge :\text{CASE} \sim \textit{nom})$ – can be informally represented as $\begin{bmatrix} \textit{noun} \\ \text{CASE}\ \ \textit{nom} \end{bmatrix}$.

(6) a. Bu ilaç *yemek-ler-den önce* al-ın-malı.
this medicine meal-PL-ABL before take-PASS-NECESS
'This medicine must be taken before meals.'
(Göksel & Kerslake 2010: 103)

b. Adam biz-e *düşman-ca* bak-ıyor-du.
man we-DAT hostile-ADVZ look-PRES.PROG-PST
'(The) man was looking at us with hostility.'
(Göksel & Kerslake 2010: 83)

c. *Son hafta-lar-da* çok yağmur yağ-dı.
last week-PL-LOC a lot rain fall-PST
'It has rained a lot in recent weeks.'
(Göksel & Kerslake 2010: 51)

d. Pamuk-lu çarşaf-lar-ı [yumuşak deterjan-la ve soğuk
cotton-ADJZ sheet-PL-ACC soft detergent-INS and cold
su-da] yıka-yın.
water-LOC wash-2P.IMP
'Wash the cotton sheets with mild detergent and in cold water.'
(Turkish Web 2012)

As for nominal modifiers, they can be 1) any PP, as in (7a); 2) any AdjP, as in (7b); or 3) an unlike category coordination where a PP is coordinated with an AdjP, as in (7c).

(7) a. *Siz-in gibi* insan-lar biz-e yardım ed-ebil-ir-ler.
you-GEN like person-PL we-DAT help do-ABIL-AOR-3PL
'People like you can help us.'
(Turkish Web 2012)

b. *Yeni* bir kitap al-dı-m.
new INDF.DET book buy-PST-1SG
'I bought a new book.'
(Göksel & Kerslake 2010: 83)

c. ... [[$_{PP}$ bir yıl boyunca] ve [$_{ADJP}$ sınır-sız]] gez-me ...
one year throughout and limit-less travel-NMZ
'... limitless sightseeing for a year ...'
(Turkish Web 2012)

Given this highly underspecified relationship between modifiers and their heads – where, for example, practically any PP can modify any verb or a noun – the relevant generalizations can be captured by the following set of constraints that directly imposes global requirements on the lexical entries of modifiers.

(8) a. $\begin{bmatrix} \textit{postp} \\ \text{MOD } \neg\textit{none} \end{bmatrix} \rightarrow \begin{bmatrix}\text{MOD}|\text{LOC}|\text{CAT}|\text{HEAD } \textit{verb} \vee \textit{noun}\end{bmatrix}$

b. $\begin{bmatrix} \textit{adj} \\ \text{MOD } \neg\textit{none} \end{bmatrix} \rightarrow \begin{bmatrix}\text{MOD}|\text{LOC}|\text{CAT}|\text{HEAD } \textit{noun}\end{bmatrix}$

c. $\begin{bmatrix} \textit{adv} \\ \text{MOD } \neg\textit{none} \end{bmatrix} \rightarrow \begin{bmatrix}\text{MOD}|\text{LOC}|\text{CAT}|\text{HEAD } \textit{verb}\end{bmatrix}$

d. $\begin{bmatrix} \textit{noun} \\ \text{CASE } \textit{loc} \vee \textit{abl} \vee \textit{ins} \\ \text{MOD } \neg\textit{none} \end{bmatrix} \rightarrow \begin{bmatrix}\text{MOD}|\text{LOC}|\text{CAT}|\text{HEAD } \textit{verb}\end{bmatrix}$

Crucially, the constraints in (8) merely specify the combinatory possibilities of modifiers but do not alone ensure that only valid instances of like and unlike coordination are licensed in adjunct positions. In any given coordinate structure, all conjuncts must specify the same MOD value and this specification must be shared with the coordination node.

For example, coordination of an AdjP and an AdvP in adjunct position is ill-formed, not because conjuncts have different categories, but because they select different heads – $[\text{MOD} \ldots \textit{noun}]$ and $[\text{MOD} \ldots \textit{verb}]$, respectively. In order to enforce this parallelism, the following constraint on *coord-phrase* is necessary:

(9) *coord-phrase* →
$\begin{bmatrix}\text{SYNSEM}|\text{HEAD} \begin{bmatrix}\text{MOD } \boxed{1} \\ \text{ARGS } \langle [\text{MOD } \boxed{1}], \ldots, [\text{MOD } \boxed{1}] \rangle \end{bmatrix}\end{bmatrix}$

(9) ensures not only that all conjuncts have the same MOD value but also that the coordination itself inherits this information. Additionally, if the conjuncts specify $[\text{MOD } \textit{none}]$ – i.e., that they are not modifiers – (9) guarantees that the coordination cannot function as a modifier as well. When combined with the *head-adjunct-phrase* constraint illustrated in (10), this analysis now licenses examples like those in (6) and (7).

(10) *head-adjunct-phrase* →
$\begin{bmatrix}\text{HD-DTR} & [\text{SYNSEM } \boxed{1}] \\ \text{NON-HD-DTRS} & \langle [\text{HEAD } [\text{MOD } \boxed{1}]] \rangle \end{bmatrix}$

(Sag 1997: 475)

4.3 Towards gradience

While the analysis presented thus far accounts for a variety of acceptable configurations of both unlike and like coordination data, it fails to take into account the finding that UCAT-LF and UCASE-LF sentences are somewhat less acceptable than their fully parallel counterparts. Since constraints are violable in Gradient HPSG, this issue can be tackled with two global constraints that would *detect* unlike category and unlike case coordination.

Accordingly, the constraint in (11) checks whether there is a categorical uniformity between the conjuncts: all members of ARGS, which are HEAD values of conjuncts (Yatabe 2004), must uniformly belong to one of the syntactic categories disjunctively specified in the constraint.

(11) *coord-phrase* →

$$\left[\text{HEAD } \boxed{1}\left[\text{ARGS } \langle \ldots \rangle\right]\right] \wedge \; [\mathtt{c}(\boxed{1}, (:\sim \textit{noun})) \vee \mathtt{c}(\boxed{1}, (:\sim \textit{adj})) \vee \mathtt{c}(\boxed{1}, (:\sim \textit{postp})) \vee \mathtt{c}(\boxed{1}, (:\sim \textit{adv})) \vee \mathtt{c}(\boxed{1}, (:\sim \textit{verb}))]$$

Unlike case coordination can be detected with the constraint in (12), which forces nominal conjuncts to bear the same case only when all the conjuncts are NPs.[9]

(12) *coord-phrase* →

$$[\left[\text{HEAD } \boxed{1}\left[\text{ARGS } \langle \ldots \rangle\right]\right] \wedge \mathtt{c}(\boxed{1}, (:\sim \textit{noun}))] \rightarrow [\mathtt{c}(\boxed{1}, (:\text{CASE} \sim \textit{nom})) \vee \mathtt{c}(\boxed{1}, (:\text{CASE} \sim \textit{gen})) \vee \mathtt{c}(\boxed{1}, (:\text{CASE} \sim \textit{acc})) \vee \mathtt{c}(\boxed{1}, (:\text{CASE} \sim \textit{dat})) \vee \mathtt{c}(\boxed{1}, (:\text{CASE} \sim \textit{loc})) \vee \mathtt{c}(\boxed{1}, (:\text{CASE} \sim \textit{abl})) \vee \mathtt{c}(\boxed{1}, (:\text{CASE} \sim \textit{ins}))]$$

Marking sentences with these two uniformity constraints for unlike category/case coordination is crucial for obtaining a modelness value that reflects the slightly reduced grammaticality of unlike coordination. However, we still need to establish constraint weights to complete the analysis and obtain modelness values.

4.4 Weight assignment

Assigning weights to specific grammar constraints requires an assumption that each experimental condition corresponds to some grammar constraint. Once

[9]The constraint can potentially be extended to cover configurations where multiple NPs are coordinated with a different syntactic category (e.g., [NP_1, NP_2 & PP]). However, since such configurations were not tested in the experiment, this extension would lack an empirical motivation.

this assumption is made, the quantified impact of an experimental condition on acceptability can be equated with the weight of its formal counterpart in the grammar.

The present analysis assumes that the relative impacts of category and case factors (i.e., LCAT- vs. UCAT- and LCASE- vs. UCASE-) correspond to the weights of the categorial uniformity constraint (see (11)) and the case uniformity constraint (see (12)), respectively. However, the grammatical function factor (i.e., -UF vs. -LF) presents a complex challenge.

Ideally, the grammatical function factor should correspond to a single constraint, much like the one-to-one correspondence between the category and case factors and their respective uniformity constraints. This single constraint would have direct access to the grammatical function of each conjunct and check for functional parallelism between them in a manner analogous to the uniformity constraints.

In the case of uniformity constraints, these checks are relatively straightforward, as the relevant features (i.e., syntactic category and case) are explicitly encoded in the HEAD values of conjuncts. In contrast, grammatical function is a more complex feature which is only implicitly (and partially) encoded in HPSG. Therefore, it is not clear how to formulate a single HPSG constraint that would enforce functional uniformity between conjuncts without revamping core HPSG assumptions. As such, no formal proposal for the functional uniformity constraint is provided in this work.[10]

In summary, the assumed correspondence between conditions and constraints is outlined in Table 1 below.

Conditions	FUNCTIONAL UNIF.	CATEGORICAL UNIF.	CASE UNIF.
LCAT-LF			
UCAT-LF		*	
LCAT-UF	*		
UCAT-UF	*	*	
LCASE-LF			
UCASE-LF			*
UCASE-UF	*		*

Table 1: Summary of the condition-constraint correspondence assumed in the present analysis. '*' indicates a violation of the corresponding constraint.

[10]Alternatively, one could consider a one-to-many correspondence between the relative impact of grammatical function factor and the lexical entries of relevant predicates, such as *sür-* 'last/continue'. However, this approach would not only significantly complicate the analysis and undermine its generalizability, but it would also suggest that the severity of selectional requirement violations (possibly) varies by predicate. While this controversial claim might be true, the relevant experiment does not deal with this question.

As for determining the numeric impacts of experimental conditions, Gradient HPSG does not make an assumption regarding the statistical model utilized for this purpose. However, for methodological soundness, the chosen model must be compatible with a repeated measures design where a participant is tested on a condition multiple times, and, accordingly, consider the dependence between observations. The present analysis relies on linear mixed-effects models to extract weights as such models can take into account the dependence between observations and the individual variability between participants and target sentences.[11]

On the basis of the aforementioned condition-constraint correspondence, a linear-mixed effects model[12] was fitted on the experimental data in question. As standard in experimental syntax, the model treated participants and items as random effects with the sentence type being the sole fixed effect.

According to the fitted model (see Table 2), a sentence that adheres to all constraints (i.e., LCAT-LF and LCASE-LF sentences) is predicted to have an average acceptability of 2.29 (on a scale from −3 to 3), as indicated by the model's intercept. Violating the functional uniformity constraint (denoted by `func_uniformity`) results in an average drop of −2.53 in acceptability, significantly exceeding the individual impacts of categorical and case mismatches, which are −0.33 and −0.62, respectively. In conclusion, these coefficients are assigned to their respective constraints in the grammar.

Fixed Effects	**Coefficients**	**Std. Error**
`(Intercept)`	2.29	0.13
`func_uniformity`	−2.53	0.24
`cat_uniformity`	−0.33	0.11
`case_uniformity`	−0.62	0.14

Table 2: Summary of the fixed effects in the fitted linear mixed-effects model

4.5 Predictions

With the relevant constraints and their weights established, modelness values can now be computed. Consider (13) and (14), which are actual UCAT-LF and UCAT-UF sentences used in the experiment.

[11] For a recent guideline on fitting linear mixed-effects models, refer to Bates, Kliegl, et al. (2015).

[12] The model was trained in R (version 4.1.2; R Core Team 2021) using the `lme4` package (Bates, Mächler, et al. 2015).

(13)	Bu	isyanlar	[[PP yıl-lar	boyunca]	ve	[NP her	gün]]
	this	rebellion-PL	year-PL	throughout	and	every	day

sür-dü.
last-PST
'These rebellions lasted for years and every day.'

(14)	* [[NP Bu	savaş-lar]	ve	[NP toprak-lar-ımız-da]]	yıl-lar-ca
	this	war-PL	and	land-PL-1PL.POSS-LOC	year-PL-ADVZ

sür-dü.
last-PST
'These wars and in our lands lasted for years.'

Table 3 illustrates both the weights (i.e., coefficients extracted from the mixed-effects model) and the modelness values for (13) and (14) based on these weights.

	FUNCTIONAL UNIF. $w = 2.53$	CATEGORICAL UNIF. $w = 0.33$	CASE UNIF. $w = 0.62$	***M***
(13)	0	1	0	-0.33
(14)	1	0	1	-3.15

Table 3: Modelness of (13) and (14). Numeric values under each constraint column indicate the number of violations per sentence, which is not greater than 1 since the sentences contain no more than one coordinate structure.

The sentence (13) violates only the categorical uniformity constraint, as the conjuncts individually satisfy the disjunctive requirements of *sür-* but bear different categories. Since not all the conjuncts are nominal, the case uniformity constraint is trivially satisfied as well. Thus, the prediction for the modelness degree of (13) is close to 0, which makes it a nearly perfect model of the grammar.

In contrast, the prediction for sentence (14), -3.15, is considerably more negative as (14) violates both functional uniformity and case uniformity constraints because the coordination that occupies the subject position involves a nominative NP (the subject) and a locative NP (an adjunct).

Modelness values can alternatively be interpreted on the original experimental scale by subtracting the non-negated modelness values from the intercept of the mixed-effects model.[13] For instance, (13) is predicted to have an acceptability score of 1.96 ($2.29 - 0.33$; INTERCEPT − MODELNESS) on

[13]This method does not apply to all statistical models and would not yield the desired results if the original scores are transformed (e.g., into z-scores).

the original scale, while (14) is predicted to score −0.86. These predictions are quite close to the actual mean scores of 1.73 and −0.62 that these sentences received in the experiment.

5 Conclusion

The picture of grammaticality derived from controlled acceptability judgment experiments is inherently gradient, a characteristic also observed in the experiment outlined in this study. To formally analyze the current experimental data within a binary framework of grammar, one would need to posit arbitrary generalizations.

In the Gradient HPSG analysis elucidated here, no such arbitrary measures were needed as the relevant observations could be modeled directly from the experimental data. Consequently, Gradient HPSG presents linguists with a promising avenue to faithfully model their experimental data by utilizing rich representations intrinsic to HPSG.

References

Baisa, Vit & Vit Suchomel. 2012. Large corpora for Turkic languages and unsupervised morphological analysis. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk & Stelios Piperidis (eds.), *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, 28–32. Istanbul, Turkey: European Language Resources Association (ELRA).

Bates, Douglas, Reinhold Kliegl, Shravan Vasishth & Harald Baayen. 2015. Parsimonious mixed models. *arXiv*. DOI: 10.48550/arxiv.1506.04967.

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48. DOI: 10.18637/jss.v067.i01.

Beavers, John & Ivan A. Sag. 2004. Coordinate ellipsis and apparent non-constituent coordination. In Stefan Müller (ed.), *Proceedings of the 11th International Conference on Head-Driven Phrase Structure Grammar*, 48–69. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2004.3.

Bolinger, D. 1961. *Generality, gradience, and the all-or-none*. The Hague: Mouton & Co.

Brew, Chris. 1995. Stochastic HPSG. In Steven Abney & Erhard Hinrichs (eds.), *Proceedings of the Seventh Conference on European Chapter of the Association for Computational Linguistics*, 83–89. San Francisco, CA: Morgan Kaufmann Publishers Inc. DOI: 10.3115/976973.976986.

Bruening, Benjamin & Eman Al Khalaf. 2020. Category mismatches in coordination revisited. *Linguistic Inquiry* 51(1). 1–36. DOI: 10.1162/ling_a_00336.

Chomsky, Noam. 1957. *Syntactic structures*. The Hague: Mouton & Co.
Chomsky, Noam. 1961. Some methodological remarks on Generative Grammar. *Word* 17(2). 219–239. DOI: 10.1080/00437956.1961.11659755.
Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, MA: The MIT Press.
Cowart, Wayne. 1997. *Experimental syntax: Applying objective methods to sentence judgments*. London: SAGE Publications.
Dalrymple, Mary. 2017. Unlike phrase structure category coordination. *Bergen Language and Linguistics Studies* 8(1). DOI: 10.15845/bells.v8i1.1332.
Featherston, Sam. 2005a. The Decathlon Model of empirical syntax. In Stephan Kepser & Marga Reis (eds.), *Linguistic evidence: Empirical, theoretical and computational perspectives* (Studies in Generative Grammar [SGG] 85), 187–208. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110197549.187.
Featherston, Sam. 2005b. Universals and grammaticality: *wh*-constraints in German and English. *Linguistics* 43(4). 667–711. DOI: 10.1515/ling.2005.43.4.667.
Göksel, Aslı & Celia Kerslake. 2010. *Turkish: An essential grammar* (Routledge Essential Grammars). London: Routledge.
Haegeman, Liliane, Ángel L. Jiménez-Fernández & Andrew Radford. 2014. Deconstructing the subject condition in terms of cumulative constraint violation. *The Linguistic Review* 31(1). 73–150. DOI: https://doi.org/10.1515/tlr-2013-0022.
Hofmeister, Philip, Laura Casasanto Staum & Ivan A. Sag. 2014. Processing effects in linguistic judgment data: (super-)additivity and reading span scores. *Language and Cognition* 6(1). 111–145. DOI: 10.1017/langcog.2013.7.
Keller, Frank. 2000. *Gradience in grammar: Experimental and computational aspects of degrees of grammaticality*. University of Edinburgh. (Ph.D. dissertation).
Keller, Frank & Theodora Alexopoulou. 2001. Phonology competes with syntax: Experimental evidence for the interaction of word order and accent placement in the realization of information structure. *Cognition* 79(3). 301–372. DOI: 10.1016/s0010-0277(00)00131-1.
Keller, Frank & Ash Asudeh. 2002. Probabilistic learning algorithms and Optimality Theory. *Linguistic Inquiry* 33(2). 225–244. DOI: 10.1162/002438902317406704.
King, Paul John. 1999. Towards truth in Head-Driven Phrase Structure Grammar. In Valia Kordoni (ed.), *Tübingen studies in Head-Driven Phrase Structure Grammar*, 301–352. Universität Tübingen. http://www.sfs.uni-tuebingen.de/sfb/reports/berichte/132/king/king.dvi.ps (26 November, 2024).

Legendre, Géraldine, Yoshiro Miyata & Paul Smolensky. 1990. *Harmonic Grammar: A formal multi-level connectionist theory of linguistic well-formedness: Theoretical foundations.* Technical Report CU-CS-465-90. University of Colorado at Boulder, Department of Computer Science.

Müller, Stefan, Anne Abeillé, Robert D. Borsley & Jean-Pierre Koenig (eds.). 2024. *Head-Driven Phrase Structure Grammar: The handbook, Second revised edition* (Empirically Oriented Theoretical Morphology and Syntax 9). Berlin: Language Science Press. DOI: 10.5281/zenodo.13637708.

Parrot, Jeffrey K. 2009. Danish vestigial case and the acquisition of vocabulary in Distributed Morphology. *Biolinguistics* 3(2). 270–302.

Patejuk, Agnieszka. 2015. *Unlike coordination in Polish: An LFG account.* Institute of Polish Language, Polish Academy of Sciences. (Ph.D. dissertation).

Patejuk, Agnieszka & Adam Przepiórkowski. 2023. Category mismatches in coordination vindicated. *Linguistic Inquiry* 54(2). 326–349. DOI: 10.1162/ling_a_00438.

Pollard, Carl & Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar* (Studies in Contemporary Linguistics). Chicago, IL: The University of Chicago Press.

Przepiórkowski, Adam. 2021. Three improvements to the HPSG model theory. In Stefan Müller & Nurit Melnik (eds.), *Proceedings of the 28th International Conference on Head-Driven Phrase Structure Grammar*, 165–185. Frankfurt am Main: Frankfurt/Main University Library. DOI: 10.21248/hpsg.2021.9.

Przepiórkowski, Adam. 2022. Coordination of unlike grammatical cases (and unlike categories). *Language* 98(3). 592–634. DOI: 10.1353/lan.0.0272.

Pullum, Geoffrey K. & Barbara C. Scholz. 2001. On the distinction between model-theoretic and generative-enumerative syntactic frameworks. In Philippe de Groote, Glyn Morrill & Christian Retoré (eds.), *Logical aspects of computational linguistics*, 17–43. Berlin: Springer.

R Core Team. 2021. *R: A language and environment for statistical computing.* R Project for statistical computing. https://www.R-project.org/.

Richter, Frank. 2004. *A mathematical formalism for linguistic theories with an application in Head-Driven Phrase Structure Grammar.* Eberhard Karls Universität Tübingen. (Ph.D. dissertation).

Richter, Frank. 2007. Closer to the truth: A new model theory for HPSG. In James Rogers & Stephan Kepser (eds.), *Model-theoretic syntax at 10*, 101–110.

Sag, Ivan A. 1997. English relative clause constructions. *Journal of Linguistics* 33(2). 431–483. DOI: 10.1017/S002222679700652X.

Sag, Ivan A., Gerald Gazdar, Thomas Wasow & Steven Weisler. 1985. Coordination and how to distinguish categories. *Natural Language & Linguistic Theory* 3(2). 117–171. http://www.jstor.org/stable/4047644 (26 November, 2024).

Sag, Ivan A. & Thomas Wasow. 2011. Performance-compatible competence grammar. In Robert Borsley & Kersti Börjars (eds.), *Non-transformational syntax*, 359–377. Chichester: Wiley-Blackwell. DOI: 10.1002/9781444395037.ch10.

Şenşekerci, Berke & Adam Przepiórkowski. 2024. Coordination of unlikes in Turkish. In Miriam Butt, Jamie Y. Findlay & Ida Toivonen (eds.), *The Proceedings of the LFG'24 Conference*, Under review. Konstanz: PubliKon.

Sorace, Antonella & Frank Keller. 2005. Gradience in linguistic data. *Lingua* 115(11). 1497–1524. DOI: 10.1016/j.lingua.2004.07.002.

Wasow, Thomas. 2024. Processing. In Stefan Müller, Anne Abeillé, Robert D. Borsley & Jean–Pierre Koenig (eds.), *Head-Driven Phrase Structure Grammar: The handbook, Second revised edition* (Empirically Oriented Theoretical Morphology and Syntax 9), 1153–1180. Berlin: Language Science Press. DOI: 10.5281/zenodo.13644940.

Weisser, Philipp. 2020. On the symmetry of case in conjunction. *Syntax* 23(1). 42–77. DOI: 10.1111/synt.12188.

Williams, Edwin. 1981. Transformationless grammar. *Linguistic Inquiry* 12. 645–653.

Yatabe, Shûichi. 2004. A comprehensive theory of coordination of unlikes. In Stefan Müller (ed.), *Proceedings of the 11th International Conference on Head-Driven Phrase Structure Grammar*, 335–355. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2004.19.