Gradient HPSG

Berke Şenşekerci

Faculty of Philosophy (Cognitive Science)
University of Warsaw

July 9, 2024

Overview

Preliminaries

1. The objective of linguistic theories is to model mental grammar.
2. To access this mental grammar, linguists often conduct acceptability judgment experiments.

Acceptability judgment experiment

How natural/acceptable do these sentences sound to you?

Sentence	1	2	3	4	5
The waitress doesn't like him.					
Waitress the doesn't like him.					

⇓

Sentence	**Mean**	**Median**
The waitress doesn't like him.	4.5	5
Waitress the doesn't like him.	1.7	2

⇓

Theory: In English NPs, determiners precede their nouns.

Problem

Problem: Transfer from experimental data to theory

1. Most grammar frameworks assume that grammaticality is binary.
2. Judgment data rarely exhibits a binary division (Keller 2000; Featherston 2005).
3. Either data is wrong or binary grammaticality assumption is wrong.

Option 1: *Data is confounded*

1. Judgment response = grammar + performance.
2. Gradience most likely stems from performance confounds.
3. Hence, some idealization is warranted.

Option 2: *Binary grammaticality assumption is too strong*

1. Judgment response = grammar + performance.
2. Gradience persists in highly controlled experiments.
3. Hence, grammaticality is most likely a gradient notion.

Adopting Solution 1: Idealization

Where to divide?

Assumption: Data is confounded. Some idealization is warranted.

Data (hypothetical):

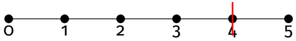
1. *The waitress doesn't like him.* (1–5; Mean: 4.5)
2. *The waitress doesn't like he.* (1–5; Mean: 2.6)
3. *Waitress the doesn't like him.* (1–5; Mean: 1.7)

How to idealize such data?

Idealization 1: Divide in the middle

- 3.1 *The waitress doesn't like him.*
- 3.2 *The waitress doesn't like he.*
- 3.3 *Waitress the doesn't like him.*

Idealization 2: 4 and above only



- 3.4 *The waitress doesn't like him.*
- 3.5 *The waitress doesn't like he.*
- 3.6 *Waitress the doesn't like him.*

Adopting Solution 1: Idealization contd.

Theoretical issues

- Different grammars from same data

$\mathcal{G}_1$ (cutoff: 2.5):

(3.1) *The waitress doesn't like him.*

(3.2) *The waitress doesn't like he.*

(3.3) *Waitress the doesn't like him.*

$\mathcal{G}_2$ (cutoff: 4):

(3.4) *The waitress doesn't like him.*

(3.5) *The waitress doesn't like he.*

(3.6) *Waitress the doesn't like him.*

- 'Judgment response = grammar + performance' does not warrant idealization:
  - Carefully designed experiments minimize the impact of performance-related confounds.
  - Performance-related confounds in AJTs are not well-understood.

Ultimately, idealizing data is riddled with problems and lacks empirical justification.

Adopting Solution 2: Gradient grammaticality

Lack of a suitable framework

Assumption: Data is reliable. Idealization is NOT warranted.

Data (hypothetical):

1. *The waitress doesn't like him.* (1–5; Mean: 4.5)
2. *The waitress doesn't like he.* (1–5; Mean: 2.6)
3. *Waitress the doesn't like him.* (1–5; Mean: 1.7)

How to analyze such data?

1. Probabilistic grammars (e.g. Brew 1995; Riezler 1999; Miyao and Tsujii 2008)
   - Training algorithms assume corpora.
   - Acceptability judgment response $\neq$ frequency (Featherston 2005; Kempen and Harbusch 2008).
2. Genuinely gradient frameworks (e.g., *Harmonic Grammar*, Legendre et al. 1990; *Linear-OT*, Keller 2000)
   - Not well-formalized
   - OT is not as rich as well-developed frameworks such as HPSG and LFG.

My proposal: A gradient version of HPSG (*work-in-progress*).

What we know about gradience

Cumulativity and weights

Acceptability of a sentence is a function of (Keller 2000; Featherston 2005):

1. The number of constraint violations (and not satisfactions)
2. The relative severity of the violated constraints

Importantly, the violation weights combine linearly (Keller 2000; Hofmeister et al. 2014).

Example

1. *The waitress doesn't like him*. (avg. acceptability = n)
2. *The waitress doesn't like he*. (avg. acceptability = $n - 1.10$)
3. *Waitress the doesn't like him*. (avg. acceptability = $n - 1.90$)
4. *Waitress the doesn't like he*. (avg. acceptability $\approxeq n - (1.90 + 1.10)$)
5. *Waitress the doesn't like boss the*. (avg. acceptability $\approxeq n - (1.90 + 1.90)$)

What we know about gradience contd.

Gradient frameworks & Harmony function

These empirical observations led to the development of frameworks (e.g., *Linear Optimality Theory*) where:

1. Constraints are violable.
2. Constraints are attached numeric weights.
3. Grammaticality is the output of a real-valued mathematical function (i.e., *Harmony function*, Legendre et al. 1990; Keller 2000).

Linear Optimality Theory (Keller 2000)

Grammar is a tuple $\langle C, w \rangle$ where:

- C is the constraint set, i.e., $C = \{C_1, C_2, ..., C_n\}$
- $w(C_i)$ is a function that maps a constraint $C_i \in C$ on its weight w_i.

Grammaticality of a sentence S is computed with *harmony function*:

$$H(S) = -\sum_i w(C_i)v(S, C_i) \tag{1}$$

$w(C_i)$ = weight of ith constraint in grammar
$v(S, C_i)$ = violation profile of ith constraint in grammar w.r.t. to sentence S

Towards gradience in HPSG

Motivations

In Linear-OT, it is formally not clear how $v(S, C_i)$ detects C_i violations in S.

1. Constraint formulations lack precision:

   CASE-MARKING: DPs must be case marked.
   SUBJ: Clauses must have subjects.

   (Grimshaw 1997)

2. Representation of S is formally impoverished.
3. **In conclusion**, there is a formal gap between constraints and S.

Why HPSG backbone is better:

1. Solid model-theoretic foundations
2. Constraints are precisely defined in a formal language.
3. S has a conspicuous formal structure.
4. **In conclusion**, there is a formally established relationship between constraints and S.

Towards gradience in HPSG

New RSRL (Richter 2004) definitions

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:
$\theta = \{\langle \delta, w \rangle \mid \delta \in \mathcal{D}_0^{\Sigma} \land w \in \mathbb{R}^+\}$

In simpler terms: Each constraint is annotated with a positive real number.

Definition 2. *Model*

For each grammar $\Gamma = \langle \Sigma, \theta \rangle$, for each Σ interpretation I = $\langle$U, **r**, S, A, R $\rangle$,
The modelness degree of I with respect to Γ is:

$$M(I) = -\sum_{i=1} |U \backslash D_I(\delta_i)| \cdot w_i$$

In simpler terms: Grammaticality is tied to *harmony function* that operates on HPSG models, which are rooted and non-exhaustive (Przepiórkowski 2021).

Towards gradience in HPSG

Demonstrating *M(I)*

Example:

$\theta = \{..., [(:\sim word) \rightarrow (:\text{PHON} \sim elephant), 0.60], ...\}$

$$I: \begin{bmatrix} headed\text{-}ph_1 & \\ \text{DTRS} & \left\langle \begin{bmatrix} word_2 & \\ \text{PHON} & the_3 \\ \text{SYNSEM} & det_4 \end{bmatrix}, \boxed{1} \begin{bmatrix} word_5 & \\ \text{PHON} & dog_6 \\ \text{SYNSEM} & noun_7 \end{bmatrix} \right\rangle \\ \text{HD-DTR} & \boxed{1} \\ ... & \end{bmatrix}$$

1. $|U \backslash D_I(\delta_i)| \cdot w_i$
2. $|U \backslash D_I((:\sim word) \rightarrow (:\text{PHON} \sim elephant))| \cdot 0.60$
3. $|U \backslash ((U \backslash D_I(:\sim word)) \cup D_I(:\text{PHON} \sim elephant))| \cdot 0.60$
4. $|\{1,2,3,4,5,6,7\} \backslash ((\{1,2,3,4,5,6,7\} \backslash \{2,5\}) \cup \emptyset)| \cdot 0.60$
5. $|\{1,2,3,4,5,6,7\} \backslash (\{1,3,4,5,7\} \cup \emptyset)| \cdot 0.60$
6. $|\{1,2,3,4,5,6,7\} \backslash \{1,3,4,5,7\}| \cdot 0.60$
7. $|\{2,5\}| \cdot 0.60$
8. $2 \cdot 0.60$
9. 1.20 w.r.t. $(:\sim word) \rightarrow (:\text{PHON} \sim elephant)$

Background

Coordination of unlikes

One widely assumed position contends that conjuncts in a given coordinate structure **must** have:

- The same syntactic category (Chomsky 1957; Williams 1981; Bruening and Khalaf 2020)
- The same case (Weisser 2020)

Counter-examples:

1. We all believe [[$_{PP}$ in positive energy] and [$_{CP}$ that what you give comes back]].

   (Patejuk and Przepiórkowski 2023)

2. [[$_{NP(acc)}$ Him] and [$_{NP(nom)}$ I]] are fighting.

   (Parrot 2009)

Controversy continues:

- Numerous attested counter-examples from Polish and English (see Przepiórkowski 2022; Patejuk and Przepiórkowski 2023)
- Lack of extensive cross-linguistic data
- Lack of experimental data

Outline

General Hypothesis

There is no universal requirement imposed by coordination that conjuncts must match both in their category and case.

So long as the conjuncts serve the same grammatical function in the sentence, valid category or case mismatches can occur in **Turkish**.

Design: Unlike category block & Unlike case block
Variables: CATEGORY/CASE MATCH × FUNCTION MATCH
Method:
- Token-set methodology (Cowart 1997).
- Likert scale (–3 to 3).

Materials: 84 target sentences (split into 4 sub-surveys) and 22 fillers.
Sample: 48 native speakers of Turkish (Mean age = 30.25)

Unlike category block

Conditions: CATEGORY (like/unlike) × FUNCTION (like/unlike)
Prediction: LCAT-LF and UCAT-LF sentences will receive higher scores compared to LCAT-UF and UCAT-UF.
Materials:
- 12 × 4 = 48 target sentences split into 4 sub-surveys
- UCAT-LF: Either unlike adjuncts or arguments

(TS1) ⟨LCAT-**LF**⟩. Bu [[$_{NP(SUBJ)}$ savaş-lar] ve [$_{NP(SUBJ)}$ isyan-lar]] toprak-lar-ımız-da yıl-lar-ca sür-dü.
this war-PL and rebellion-PL land-PL-1PL.POSS-LOC year-PL-ADVZ continue-PST

'These wars and rebellions continued for years in our lands.'

⟨UCAT-**LF**⟩. Bu isyan-lar [[$_{PP(OBL)}$ yıl-lar boyunca] ve [$_{NP(OBL)}$ her gün]] sür-dü.
this rebellion-PL year-PL throughout and every day continue-PST

'These rebellions continued for years and every day.'

Unlike category block contd.

Conditions: CATEGORY (like/unlike) × FUNCTION (like/unlike)
Prediction: LCAT-LF and UCAT-LF sentences will receive higher scores compared to LCAT-UF and UCAT-UF.
Materials:
- 12 × 4 = 48 target sentences split into 4 sub-surveys.
- UCAT-LF: Either unlike adjuncts or arguments

(TS1) ⟨LCAT-**UF**⟩. [[$_{NP(SUBJ)}$ Bu savaş-lar] ve [$_{NP(ADJUNCT)}$ toprak-lar-ımız-da]] yıl-lar-ca sür-dü.
this war-PL and land-PL-1PL.POSS-LOC year-PL-ADVZ continue-PST

Lit. '*These wars and in our lands lasted for years.*'

⟨UCAT-**UF**⟩. [[$_{NP(SUBJ)}$ Bu savaş-lar] ve [$_{AdvP(ADJUNCT)}$ yıl-lar-ca]] sür-dü.
this war-PL and year-PL-ADVZ continue-PST

Lit. '*These wars and for years lasted.*'

Unlike case block

Conditions: LCASE-LF, UCASE-LF, LCASE-UF, UCASE-UF (due to Turkish case system)
Prediction: LCASE-LF and UCASE-LF sentences will receive higher scores compared to LCASE-UF and UCASE-UF.
Materials:
- $12 \times 3 = 36$ target sentences split into 4 sub-surveys.
- UCASE-LF: Unlike adjuncts

(TS4) ⟨LCASE-**LF**⟩. Oğlum-a internet-ten [[$_{\text{NOM(OBJ)}}$ şapka] ve
son-DAT internet-ABL hat and
[$_{\text{NOM(OBJ)}}$ ayakkabı]] al-dı-m
shoe buy-PST-1SG

'I bought my son a pair of shoes and a hat through the internet.'

⟨UCASE-**LF**⟩. Oğlum-a [[$_{\text{ABL(ADJUNCT)}}$ internet-ten] ve
son-DAT internet-ABL and
[$_{\text{INS(ADJUNCT)}}$ kredi kart-ı-yla]] ayakkabı al-dı-m
credit card-3P-INS shoe buy-PST-1SG

'I bought my son a pair of shoes through the internet and by credit card.'

Unlike case block contd.

Conditions: LCASE-LF, UCASE-LF, UCASE-UF

Prediction: LCASE-LF and UCASE-LF sentences will receive higher scores compared to LCASE-UF and UCASE-UF.

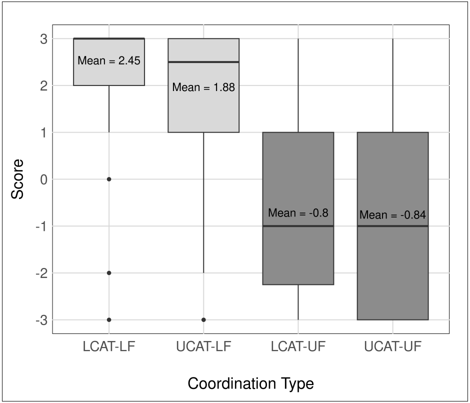
Materials:
- $12 \times 3 = 36$ target sentences split into 4 sub-surveys.
- UCASE-LF: Unlike adjuncts

(TS4) ⟨UCASE-**UF**⟩. [[$_{\text{DAT(OBL)}}$ Oğlum-a] ve [$_{\text{ACC(OBJ)}}$ iste-diğ-i ayakkabı-yı]] internet-ten al-dı-m

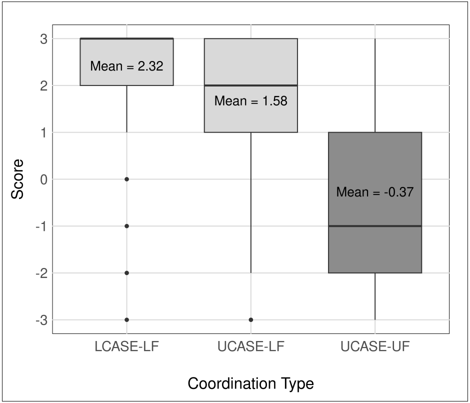
Oğlum-a]	ve	iste-diğ-i	ayakkabı-yı]]	internet-ten	al-dı-m
son-DAT	and	want-PTCP-3P	shoe-ACC	internet-ABL	buy-PST-1SG

Lit. '*I bought my son and the pair of shoes he wanted through the internet.*'

Results



(a) Unlike category coordination block



(b) Unlike case coordination block

- In both blocks, X-LF conditions received significantly higher scores.
- In unlike case block, UCASE-LF is acceptable but significantly lower than LCASE-LF ($p < 0.001$)

Constraints
Adjuncts

Idea: Model both unlike argument and unlike adjunct configurations in terms of disjunctive selectional requirements (Yatabe 2004; Przepiórkowski 2021)

Modifiers: Rather underspecified relationship between modifiers and their heads

$$\begin{bmatrix} \textit{postp} \\ \text{MOD} \;\; \neg\textit{none} \end{bmatrix} \rightarrow \begin{bmatrix} \text{MOD}|\text{LOC}|\text{CAT}|\text{HEAD} \;\; \textit{verb} \vee \textit{noun} \end{bmatrix} \qquad \begin{bmatrix} \textit{adj} \\ \text{MOD} \;\; \neg\textit{none} \end{bmatrix} \rightarrow \begin{bmatrix} \text{MOD}|\text{LOC}|\text{CAT}|\text{HEAD} \;\; \textit{noun} \end{bmatrix}$$

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

The compatibility is, in turn, checked by *head-adjunct-phrase* constraint (Sag 1997)

$$\textit{head-adjunct-phrase} \rightarrow \begin{bmatrix} \text{HD-DTR} & \begin{bmatrix} \text{SYNSEM} \;\; \boxed{1} \end{bmatrix} \\ \text{NON-HD-DTRS} & \left\langle \begin{bmatrix} \text{HEAD} \;\; \begin{bmatrix} \text{MOD} \;\; \boxed{1} \end{bmatrix} \end{bmatrix} \right\rangle \end{bmatrix}$$

Constraints

Arguments

Arguments: Disjunctive requirements imposed by the predicate on its complements (via relation c and second-order HPSG).

LE for *sür-* ‘to last/continue’:

$$\begin{bmatrix} \text{PHON} & \langle \textit{SÜR} \rangle \\ \text{SYNSEM}|\text{CAT}|\text{VALENCE} & \begin{bmatrix} \text{SUBJ} & \left\langle \begin{bmatrix} \text{CAT}|\text{HEAD}\ \boxed{1} \end{bmatrix} \right\rangle \\ \text{COMPS} & \left\langle \begin{bmatrix} \text{CAT}|\text{HEAD}\ \boxed{2} \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix} \begin{array}{l} \wedge\ \alpha_1 \approx (:\sim \textit{noun} \wedge :\text{CASE} \sim \textit{nom}) \\ \wedge\ \alpha_2 \approx [(:\sim \textit{noun} \wedge :\text{CASE} \sim \textit{nom}) \vee \\ \qquad (:\sim \textit{postp} \wedge :\text{PFORM} \sim \textit{boyunca}) \vee \\ \qquad (:\sim \textit{adv})] \\ \wedge\ \text{c}(\boxed{1}, \alpha_1) \wedge \text{c}(\boxed{2}, \alpha_2) \end{array}$$

The compatibility is, in turn, checked by *head-comp-phrase* and *head-subj-phrase* constraints (Sag 1997)

head-subj-phrase $\rightarrow$

$$\begin{bmatrix} \text{SUBJ} & \langle\ \rangle \\ \text{HD-DTR} & \begin{bmatrix} \text{SUBJ} & \langle \boxed{1} \rangle \\ \text{SPR} & \langle\ \rangle \end{bmatrix} \\ \text{NON-HD-DTRS} & \left\langle \begin{bmatrix} \text{SS}\ \boxed{1} \end{bmatrix} \right\rangle \end{bmatrix}$$

head-comp-phrase $\rightarrow$

$$\begin{bmatrix} \text{COMPS} & \langle\ \rangle \\ \text{HD-DTR} & \begin{bmatrix} \text{COMPS} & \langle \boxed{1}, \ldots, \boxed{n} \rangle \end{bmatrix} \\ \text{NON-HD-DTRS} & \left\langle \begin{bmatrix} \text{SS}\ \boxed{1} \end{bmatrix}, \ldots, \begin{bmatrix} \text{SS}\ \boxed{n} \end{bmatrix} \right\rangle \end{bmatrix}$$

Constraints

Global constraints on coordination

So far, valid UCAT-LF and UCASE-LF configurations are accounted for.

However, UCAT-LF and UCASE-LF are less acceptable than their fully parallel counterparts.

Idea: Constraints that 'detect' unlike category and unlike case coordination.

CATEGORICAL UNIFORMITY CONSTRAINT

coord-phrase $\rightarrow$

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

CASE UNIFORMITY CONSTRAINT

coord-phrase $\rightarrow$

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

Adding weights

Gradient HPSG is **agnostic towards** the method of extracting weights from experimental data:

- Crude analysis (e.g., differences between means)
- Simple/multiple linear regression
- ML models (e.g., support vector machine, random forest)
- Linear mixed effects models

Constraints	Estimated weight
head-x-phrase	−2.40
cat-uniformity	−0.67
case-uniformity	−0.66

Table: LMEM trained on the experimental data

Predictions

Reconsider:

(a.) Bu isyan-lar [[$_{PP(COMP)}$ yıl-lar boyunca] ve [$_{NP(COMP)}$ her gün]] sür-dü.
this rebellion-PL year-PL throughout and every day continue-PST

'*These rebellions continued for years and every day.*'

(b.) [[$_{NP.nom(SUBJ)}$ Bu savaş-lar] ve [$_{NP.loc(MOD)}$ toprak-lar-ımız-da]] yıl-lar-ca sür-dü.
this war-PL and land-PL-1PL.POSS-LOC year-PL-ADVZ continue-PST

Lit. '*These wars and in our lands lasted for years.*'

	head-x-phrase $w = 2.40$	cat-uniformity $w = 0.67$	case-uniformity $w = 0.66$	M(I)	Prediction	Actual (Mean)
(a.)	0	1	0	-0.67	**1.72**	**1.73**
(b.)	1	0	1	-3.06	**-0.67**	**-0.62**

Table: Modelness & Prediction & Actual

Summary

There is a clash between experimental data and theories of grammar:

- Data is confounded; idealization is warranted.
- Data is reliable; gradience should be modeled.

Existing options are NOT good enough:

- Not well-formalized
- Not intricate enough
- Not compatible with experimental data

Gradient HPSG is an amalgamation of HPSG and Linear-OT:

- Grammar is type hierarchy & weighted constraints.
- Grammaticality is a real-value.
- AJT data for training

References I

Brew, C. (1995). Stochastic HPSG. In Abney, S. P. and Hinrichs, E. W., editors, *Seventh Conference of the European Chapter of the Association for Computational Linguistics*, Dublin, Ireland. Association for Computational Linguistics.

Bruening, B. and Khalaf, E. A. (2020). Category mismatches in coordination revisited. *Linguistic Inquiry*, 51(1):1–36.

Chomsky, N. (1957). *Syntactic Structures*. Mouton & Co., The Hague.

Cowart, W. (1997). *Experimental Syntax: Applying Objective Methods to Sentence Judgments*. SAGE Publications.

Featherston, S. (2005). The Decathlon Model of empirical syntax. In Kepser, S. and Reis, M., editors, *Linguistic Evidence: Empirical, Theoretical and Computational Perspectives*, pages 187–208. De Gruyter Mouton.

Grimshaw, J. B. (1997). Projection, heads, and optimality. *Linguistic Inquiry*, 28:373–422.

Hofmeister, P., Casasanto Staum, L., and Sag, I. A. (2014). Processing effects in linguistic judgment data: (super-)additivity and reading span scores. *Language and Cognition*, 6(1):111–145.

Keller, F. (2000). *Gradience in Grammar: Experimental and Computational Aspects of Degrees of Grammaticality*. Phd thesis, University of Edinburgh.

References II

Kempen, G. and Harbusch, K. (2008). Comparing linguistic judgments and corpus frequencies as windows on grammatical competence: A study of argument linearization in german clauses. In Steube, A., editor, *The Discourse Potential of Underspecified Structures*, pages 179–192. Walter de Gruyter, Berlin, New York.

Legendre, G., Miyata, Y., and Smolensky, P. (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.

Miyao, Y. and Tsujii, J. (2008). Feature forest models for probabilistic HPSG parsing. *Computational Linguistics*, 34(1):35–80.

Parrot, J. K. (2009). Danish vestigial case and the acquisition of vocabulary in distributed morphology. *Biolinguistics*, 3(2):270–302.

Patejuk, A. and Przepiórkowski, A. (2023). Category Mismatches in Coordination Vindicated. *Linguistic Inquiry*, pages 1–24.

Przepiórkowski, A. (2021). Three improvements to the hpsg model theory. In Müller, S. and Melnik, N., editors, *Proceedings of the 28th International Conference on Head-Driven Phrase Structure Grammar, Online (Frankfurt/Main)*, pages 165–185, Frankfurt/Main. University Library.

Przepiórkowski, A. (2022). Coordination of unlike grammatical cases (and unlike categories). *Language*, 98(3):592–634.

References III

Richter, F. (2004). *A Mathematical Formalism for Linguistic Theories with an Application in Head-Driven Phrase Structure Grammar*. Phd thesis, Eberhard Karls Universität Tübingen.

Riezler, S. (1999). *Probabilistic Constraint Logic Programming*. PhD thesis, University of Stuttgart.

Sag, I. A. (1997). English relative clause constructions. *Journal of Linguistics*, 33(2):431–483.

Weisser, P. (2020). On the symmetry of case in conjunction. *Syntax*, 23(1):42–77.

Williams, E. (1981). Transformationless grammar. *Linguistic Inquiry*, 12:645–653.

Yatabe, S. (2004). A comprehensive theory of coordination of unlikes. In Müller, S., editor, *Proceedings of the 11th International Conference on Head-Driven Phrase Structure Grammar, Center for Computational Linguistics, Katholieke Universiteit Leuven*, pages 335–355, Stanford, CA. CSLI Publications.