Gradient Selectional Restrictions in HPSG

Adam Przepiórkowski[1,2], Berke Şenşekerci[2]

[1]Polish Academy of Sciences (Institute of Computer Science)
[2]University of Warsaw (Faculty of Philosophy)

HPSG 2026

Bergen, 4th August 2026

Syntactic selectional restrictions

Some **selectional restrictions** are purely **syntactic**, e.g.:

- *Homer **waited*** [$_{\text{PP}}$ *for Marge*].
- *Homer **awaited*** [$_{\text{NP}}$ *Marge*].

Syntactic selectional restrictions are usually **assumed to be hard**, e.g.:

- **Homer **waited*** [$_{\text{NP}}$ *Marge*].
- **Homer **awaited*** [$_{\text{PP}}$ *for Marge*].

Aims of this presentation:

- present **evidence that syntactic selectional restrictions may be soft** (“gradient”),
- propose how to **model gradience in HPSG** (based on earlier work),
- especially **gradient selectional restrictions**.

Syntactic selectional restrictions

Some **selectional restrictions** are purely **syntactic**, e.g.:

- *Homer **waited** [$_{PP}$ for Marge].*
- *Homer **awaited** [$_{NP}$ Marge].*

Syntactic selectional restrictions are usually **assumed to be hard**, e.g.:

- **Homer **waited** [$_{NP}$ Marge].*
- **Homer **awaited** [$_{PP}$ for Marge].*

Aims of this presentation:

- present **evidence that syntactic selectional restrictions may be soft** ("gradient"),
- propose how to **model gradience in HPSG** (based on earlier work),
- especially **gradient selectional restrictions**.

Syntactic selectional restrictions

Some **selectional restrictions** are purely **syntactic**, e.g.:

- *Homer **waited** [$_{\text{PP}}$ for Marge].*
- *Homer **awaited** [$_{\text{NP}}$ Marge].*

Syntactic selectional restrictions are usually **assumed to be hard**, e.g.:

- **Homer **waited** [$_{\text{NP}}$ Marge].*
- **Homer **awaited** [$_{\text{PP}}$ for Marge].*

Aims of this presentation:

- present **evidence that syntactic selectional restrictions may be soft** ("gradient"),
- propose how to **model gradience in HPSG** (based on earlier work),
- especially **gradient selectional restrictions**.

Syntactic selectional restrictions

Some **selectional restrictions** are purely **syntactic**, e.g.:

- *Homer **waited** [$_{\text{PP}}$ for Marge].*
- *Homer **awaited** [$_{\text{NP}}$ Marge].*

Syntactic selectional restrictions are usually **assumed to be hard**, e.g.:

- **Homer **waited** [$_{\text{NP}}$ Marge].*
- **Homer **awaited** [$_{\text{PP}}$ for Marge].*

Aims of this presentation:

- present **evidence that syntactic selectional restrictions may be soft** ("gradient"),
- propose how to **model gradience in HPSG** (based on earlier work),
- especially **gradient selectional restrictions**.

Syntactic selectional restrictions

Some **selectional restrictions** are purely **syntactic**, e.g.:

- *Homer **waited** [$_{\text{PP}}$ for Marge].*
- *Homer **awaited** [$_{\text{NP}}$ Marge].*

Syntactic selectional restrictions are usually **assumed to be hard**, e.g.:

- **Homer **waited** [$_{\text{NP}}$ Marge].*
- **Homer **awaited** [$_{\text{PP}}$ for Marge].*

Aims of this presentation:

- present **evidence that syntactic selectional restrictions may be soft** ("gradient"),
- propose how to **model gradience in HPSG** (based on earlier work),
- especially **gradient selectional restrictions**.

Syntactic selectional restrictions

Some **selectional restrictions** are purely **syntactic**, e.g.:

- *Homer **waited** [$_{\text{PP}}$ for Marge].*
- *Homer **awaited** [$_{\text{NP}}$ Marge].*

Syntactic selectional restrictions are usually **assumed to be hard**, e.g.:

- **Homer **waited** [$_{\text{NP}}$ Marge].*
- **Homer **awaited** [$_{\text{PP}}$ for Marge].*

Aims of this presentation:

- present **evidence that syntactic selectional restrictions may be soft** ("gradient"),
- propose how to **model gradience in HPSG** (based on earlier work),
- especially **gradient selectional restrictions**.

Syntactic selectional restrictions

Some **selectional restrictions** are purely **syntactic**, e.g.:

- *Homer **waited** [$_{\text{PP}}$ for Marge].*
- *Homer **awaited** [$_{\text{NP}}$ Marge].*

Syntactic selectional restrictions are usually **assumed to be hard**, e.g.:

- *Homer **waited** [$_{\text{NP}}$ Marge].
- *Homer **awaited** [$_{\text{PP}}$ for Marge].

Aims of this presentation:

- present **evidence that syntactic selectional restrictions may be soft** ("gradient"),
- propose how to **model gradience in HPSG** (based on earlier work),
- especially **gradient selectional restrictions**.

Syntactic selectional restrictions

Some **selectional restrictions** are purely **syntactic**, e.g.:

- *Homer **waited** [$_{\text{PP}}$ for Marge].*
- *Homer **awaited** [$_{\text{NP}}$ Marge].*

Syntactic selectional restrictions are usually **assumed to be hard**, e.g.:

- **Homer **waited** [$_{\text{NP}}$ Marge].*
- **Homer **awaited** [$_{\text{PP}}$ for Marge].*

Aims of this presentation:

- present **evidence that syntactic selectional restrictions may be soft** (“gradient”),
- propose how to **model gradience in HPSG** (based on earlier work),
- especially **gradient selectional restrictions**.

Experimental Evidence 1

An **acceptability judgement experiment** (N = 70; F = 24, M = 46; Prolific, LimeSurvey, R):

- we looked at **8 predicates**: *account (for), annoyed (by), ashamed (of), depend (on), familiar (with), speak (about), suffer (from) talk (about)*;
- **for each, 8 experimental items** ("tokensets") like this (i.e., 8 × 8 = 64 altogether):
  - **PP** condition (assumed to be fully grammatical):
    *Justin Bieber can **depend** [$_{PP}$ on the fact [$_{CP}$ that his fans still love his early songs]].*
  - **CP** condition (of unclear grammaticality):
    *Justin Bieber can **depend** [$_{CP}$ that his fans still love his early songs].*

The **Thermometer Method** was employed, argued to be superior to the standard Likert Scale (Featherston 2008, 2009).

Among the fillers, there were 15 "**standard items**" – 3 sentences for each of the acceptability levels A–E (Featherston 2009, Gerbrich *et al.* 2019).

Joint work with Katarzyna Kuś (University of Warsaw) and Agnieszka Patejuk (Polish Academy of Sciences).

Experimental Evidence 1

An **acceptability judgement experiment** (N = 70; F = 24, M = 46; Prolific, LimeSurvey, R):

- we looked at **8 predicates**: *account (for), annoyed (by), ashamed (of), depend (on), familiar (with), speak (about), suffer (from) talk (about)*;
- **for each, 8 experimental items** ("tokensets") like this (i.e., $8 \times 8 = 64$ altogether):
  - **PP** condition (assumed to be fully grammatical):
    *Justin Bieber can **depend*** [$_{\text{PP}}$ *on the fact* [$_{\text{CP}}$ *that his fans still love his early songs*]].
  - **CP** condition (of unclear grammaticality):
    *Justin Bieber can **depend*** [$_{\text{CP}}$ *that his fans still love his early songs*].

The **Thermometer Method** was employed, argued to be superior to the standard Likert Scale (Featherston 2008, 2009).

Among the fillers, there were 15 "**standard items**" – 3 sentences for each of the acceptability levels A–E (Featherston 2009, Gerbrich *et al.* 2019).

Experimental Evidence 1

An **acceptability judgement experiment** (N = 70; F = 24, M = 46; Prolific, LimeSurvey, R):

- we looked at **8 predicates**: *account (for), annoyed (by), ashamed (of), depend (on), familiar (with), speak (about), suffer (from) talk (about)*;
- **for each, 8 experimental items** ("tokensets") like this (i.e., 8 × 8 = 64 altogether):
  - **PP** condition (assumed to be fully grammatical):
    *Justin Bieber can **depend*** [$_{PP}$ *on the fact* [$_{CP}$ *that his fans still love his early songs*]].
  - **CP** condition (of unclear grammaticality):
    *Justin Bieber can **depend*** [$_{CP}$ *that his fans still love his early songs*].

The **Thermometer Method** was employed, argued to be superior to the standard Likert Scale (Featherston 2008, 2009).

Among the fillers, there were 15 "**standard items**" – 3 sentences for each of the acceptability levels A–E (Featherston 2009, Gerbrich *et al.* 2019).

Experimental Evidence 1

An **acceptability judgement experiment** (N = 70; F = 24, M = 46; Prolific, LimeSurvey, R):

- we looked at **8 predicates**: *account (for), annoyed (by), ashamed (of), depend (on), familiar (with), speak (about), suffer (from) talk (about)*;
- **for each, 8 experimental items** ("tokensets") like this (i.e., 8 × 8 = 64 altogether):
  - **PP** condition (assumed to be fully grammatical):
    *Justin Bieber can **depend** [$_{PP}$ on the fact [$_{CP}$ that his fans still love his early songs]].*
  - **CP** condition (of unclear grammaticality):
    *Justin Bieber can **depend** [$_{CP}$ that his fans still love his early songs].*

The **Thermometer Method** was employed, argued to be superior to the standard Likert Scale (Featherston 2008, 2009).

Among the fillers, there were 15 "**standard items**" – 3 sentences for each of the acceptability levels A–E (Featherston 2009, Gerbrich *et al.* 2019).

Joint work with Katarzyna Kuś (University of Warsaw) and Agnieszka Patejuk (Polish Academy of Sciences).

Experimental Evidence 1

An **acceptability judgement experiment** (N = 70; F = 24, M = 46; Prolific, LimeSurvey, R):

- we looked at **8 predicates**: *account (for), annoyed (by), ashamed (of), depend (on), familiar (with), speak (about), suffer (from) talk (about)*;
- **for each**, **8 experimental items** ("tokensets") like this (i.e., $8 \times 8 = 64$ altogether):
  - **PP** condition (assumed to be fully grammatical):
    *Justin Bieber can **depend*** [$_{PP}$ *on the fact* [$_{CP}$ *that his fans still love his early songs*]].
  - CP condition (of unclear grammaticality):
    *Justin Bieber can **depend*** [$_{CP}$ *that his fans still love his early songs*].

The **Thermometer Method** was employed, argued to be superior to the standard Likert Scale (Featherston 2008, 2009).

Among the fillers, there were 15 "**standard items**" – 3 sentences for each of the acceptability levels A–E (Featherston 2009, Gerbrich *et al.* 2019).

Joint work with Katarzyna Kuś (University of Warsaw) and Agnieszka Patejuk (Polish Academy of Sciences).

Experimental Evidence 1

An **acceptability judgement experiment** (N = 70; F = 24, M = 46; Prolific, LimeSurvey, R):

- we looked at **8 predicates**: *account (for), annoyed (by), ashamed (of), depend (on), familiar (with), speak (about), suffer (from) talk (about)*;
- **for each, 8 experimental items** ("tokensets") like this (i.e., 8 × 8 = 64 altogether):
  - **PP** condition (assumed to be fully grammatical):
    *Justin Bieber can **depend** [$_{PP}$ on the fact [$_{CP}$ that his fans still love his early songs]].*
  - **CP** condition (of unclear grammaticality):
    *Justin Bieber can **depend** [$_{CP}$ that his fans still love his early songs].*

The **Thermometer Method** was employed, argued to be superior to the standard Likert Scale (Featherston 2008, 2009).

Among the fillers, there were 15 "**standard items**" – 3 sentences for each of the acceptability levels A–E (Featherston 2009, Gerbrich *et al.* 2019).

Joint work with Katarzyna Kuś (University of Warsaw) and Agnieszka Patejuk (Polish Academy of Sciences).

Experimental Evidence 1

An **acceptability judgement experiment** (N = 70; F = 24, M = 46; Prolific, LimeSurvey, R):

- we looked at **8 predicates**: *account (for), annoyed (by), ashamed (of), depend (on), familiar (with), speak (about), suffer (from) talk (about)*;
- **for each, 8 experimental items** ("tokensets") like this (i.e., 8 × 8 = 64 altogether):
  - **PP** condition (assumed to be fully grammatical):
    *Justin Bieber can **depend*** $[_{PP}$ *on the fact* $[_{CP}$ *that his fans still love his early songs*$]]$.
  - **CP** condition (of unclear grammaticality):
    *Justin Bieber can **depend*** $[_{CP}$ *that his fans still love his early songs*$]$.

The **Thermometer Method** was employed, argued to be superior to the standard Likert Scale (Featherston 2008, 2009).

Among the fillers, there were 15 "**standard items**" – 3 sentences for each of the acceptability levels A–E (Featherston 2009, Gerbrich *et al.* 2019).

Joint work with Katarzyna Kuś (University of Warsaw) and Agnieszka Patejuk (Polish Academy of Sciences).

Experimental Evidence 1

An **acceptability judgement experiment** (N = 70; F = 24, M = 46; Prolific, LimeSurvey, R):

- we looked at **8 predicates**: *account (for), annoyed (by), ashamed (of), depend (on), familiar (with), speak (about), suffer (from) talk (about)*;
- **for each**, **8 experimental items** ("tokensets") like this (i.e., 8 × 8 = 64 altogether):
  - **PP** condition (assumed to be fully grammatical):
    *Justin Bieber can **depend*** $[_{\text{PP}}$ *on the fact* $[_{\text{CP}}$ *that his fans still love his early songs*$]]$.
  - **CP** condition (of unclear grammaticality):
    *Justin Bieber can **depend*** $[_{\text{CP}}$ *that his fans still love his early songs*$]$.

The **Thermometer Method** was employed, argued to be superior to the standard Likert Scale (Featherston 2008, 2009).

Among the fillers, there were 15 "**standard items**" – 3 sentences for each of the acceptability levels A–E (Featherston 2009, Gerbrich *et al.* 2019).

kognilab Laboratory of Experimental Philosophy

Robert found but Emma lost two pound coins.

Help: There are two sentences on the scale. Their location on the scale (one at 20, and the other at 30) indicates their respective naturalness scores. Your task is to rate the naturalness of the sentence above, considering the scores of these two sentences.

20
The father fetches the wholemeal bread them.

30
The father fetches the children the wholemeal bread.

Next

Standard Items

A.
- The patient fooled the dentist by pretending to be in pain.
- There's a statue in the middle of the square.
- The winter is very harsh in the North.

B.
- Before every lesson the teacher must prepare their materials.
- Jack doesn't boast about his being elected chairman.
- John cleaned his motorbike with which cleaning cloth?

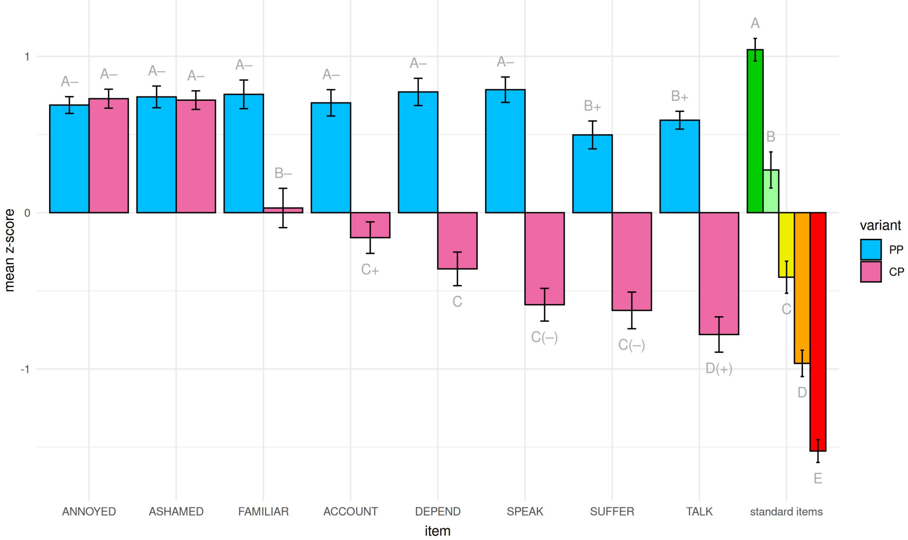
C.
- Anna loves, but Linda hates, eating popcorn at the cinema.
- Most people like very much a cup of tea in the morning.
- The striker fouled deliberately the goalkeeper.

D.
- Who did he whisper that had unfairly condemned the prisoner?
- The old fisherman took her pipe out of mouth and began story.
- Which professor did you claim that the student really admires him?

E.
- Historians wondering what cause is disappear civilization.
- Old man he work garden grow many flowers and vegetable.
- Student must read much book for they become clever.

Experimental Evidence 2

Average z-scores (with 95% confidence intervals):



Experimental Evidence 3

Summary of experimental evidence:

- the acceptability of **PP variants was always high** (between A– and B+),
- the acceptability of **CP variants varied greatly**:
  - from **fully acceptable** *annoyed* and *ashamed* (A–),
  - through **marginally acceptable** *familiar* (B–) and *account* (C+),
  - to the **least acceptable** *speak* (C(–)), *suffer* (C(–)), and *talk* (D(+));
- **PP vs. CP difference** (according to linear mixed effects models):
  - not statistically significant for *annoyed* and *ashamed*,
  - statistically significant ($p < .001$) for the other six predicates;
- all **means of CP conditions** – with the exception of *talk* – **statistically significantly higher** than that of standard items D.

Length and complexity were controlled for, so it is **unlikely that these differences reflect processing**.

This justifies positing the following **hypothesis**:
Syntactic selectional restrictions may be gradient. **Corollary**: **Grammaticality is gradient**.

Experimental Evidence 3

Summary of experimental evidence:

- the acceptability of **PP variants was always high** (between A− and B+),
- the acceptability of **CP variants varied greatly**:
  - from **fully acceptable** *annoyed* and *ashamed* (A−),
  - through **marginally acceptable** *familiar* (B−) and *account* (C+),
  - to the **least acceptable** *speak* (C(−)), *suffer* (C(−)), and *talk* (D(+));
- **PP vs. CP difference** (according to linear mixed effects models):
  - not statistically significant for *annoyed* and *ashamed*,
  - statistically significant ($p < .001$) for the other six predicates;
- all **means of CP conditions** – with the exception of *talk* – **statistically significantly higher** than that of standard items D.

Length and complexity were controlled for, so it is **unlikely that these differences reflect processing**.

This justifies positing the following **hypothesis**:
Syntactic selectional restrictions may be gradient. **Corollary**: **Grammaticality is gradient.**

Experimental Evidence 3

Summary of experimental evidence:

- the acceptability of **PP variants was always high** (between A− and B+),
- the acceptability of **CP variants varied greatly**:
  - from **fully acceptable** *annoyed* and *ashamed* (A−),
  - through **marginally acceptable** *familiar* (B−) and *account* (C+),
  - to the **least acceptable** *speak* (C(−)), *suffer* (C(−)), and *talk* (D(+));
- **PP vs. CP difference** (according to linear mixed effects models):
  - not statistically significant for *annoyed* and *ashamed*,
  - statistically significant ($p < .001$) for the other six predicates;
- all **means of CP conditions** – with the exception of *talk* – **statistically significantly higher** than that of standard items D.

Length and complexity were controlled for, so it is **unlikely that these differences reflect processing**.

This justifies positing the following **hypothesis**:
Syntactic selectional restrictions may be gradient. **Corollary**: **Grammaticality is gradient.**

Experimental Evidence 3

Summary of experimental evidence:

- the acceptability of **PP variants was always high** (between A− and B+),
- the acceptability of **CP variants varied greatly**:
  - from **fully acceptable** *annoyed* and *ashamed* (A−),
  - through **marginally acceptable** *familiar* (B−) and *account* (C+),
  - to the **least acceptable** *speak* (C(−)), *suffer* (C(−)), and *talk* (D(+));
- **PP vs. CP difference** (according to linear mixed effects models):
  - not statistically significant for *annoyed* and *ashamed*,
  - statistically significant ($p < .001$) for the other six predicates;
- all **means of CP conditions** – with the exception of *talk* – **statistically significantly higher** than that of standard items D.

Length and complexity were controlled for, so it is **unlikely that these differences reflect processing**.

This justifies positing the following **hypothesis**:
Syntactic selectional restrictions may be gradient. **Corollary**: **Grammaticality is gradient.**

Experimental Evidence 3

Summary of experimental evidence:

- the acceptability of **PP variants was always high** (between A− and B+),
- the acceptability of **CP variants varied greatly**:
  - from **fully acceptable** *annoyed* and *ashamed* (A−),
  - through **marginally acceptable** *familiar* (B−) and *account* (C+),
  - to the **least acceptable** *speak* (C(−)), *suffer* (C(−)), and *talk* (D(+));
- **PP vs. CP difference** (according to linear mixed effects models):
  - not statistically significant for *annoyed* and *ashamed*,
  - statistically significant ($p < .001$) for the other six predicates;
- all **means of CP conditions** – with the exception of *talk* – **statistically significantly higher** than that of standard items D.

Length and complexity were controlled for, so it is **unlikely that these differences reflect processing**.

This justifies positing the following **hypothesis**:
Syntactic selectional restrictions may be gradient. **Corollary**: **Grammaticality is gradient.**

Experimental Evidence 3

Summary of experimental evidence:

- the acceptability of **PP variants was always high** (between A− and B+),
- the acceptability of **CP variants varied greatly**:
  - from **fully acceptable** *annoyed* and *ashamed* (A−),
  - through **marginally acceptable** *familiar* (B−) and *account* (C+),
  - to the **least acceptable** *speak* (C(−)), *suffer* (C(−)), and *talk* (D(+));
- **PP vs. CP difference** (according to linear mixed effects models):
  - not statistically significant for *annoyed* and *ashamed*,
  - statistically significant ($p < .001$) for the other six predicates;
- all **means of CP conditions** – with the exception of *talk* – **statistically significantly higher** than that of standard items D.

Length and complexity were controlled for, so it is **unlikely that these differences reflect processing**.

This justifies positing the following **hypothesis**:
Syntactic selectional restrictions may be gradient. **Corollary**: **Grammaticality is gradient.**

Experimental Evidence 3

Summary of experimental evidence:

- the acceptability of **PP variants was always high** (between A− and B+),
- the acceptability of **CP variants varied greatly**:
  - from **fully acceptable** *annoyed* and *ashamed* (A−),
  - through **marginally acceptable** *familiar* (B−) and *account* (C+),
  - to the **least acceptable** *speak* (C(−)), *suffer* (C(−)), and *talk* (D(+));
- **PP vs. CP difference** (according to linear mixed effects models):
  - not statistically significant for *annoyed* and *ashamed*,
  - statistically significant ($p < .001$) for the other six predicates;
- all **means of CP conditions** – with the exception of *talk* – **statistically significantly higher** than that of standard items D.

Length and complexity were controlled for, so it is **unlikely that these differences reflect processing**.

This justifies positing the following **hypothesis**:
Syntactic selectional restrictions may be gradient. **Corollary**: **Grammaticality is gradient.**

Experimental Evidence 3

Summary of experimental evidence:

- the acceptability of **PP variants was always high** (between A− and B+),
- the acceptability of **CP variants varied greatly**:
  - from **fully acceptable** *annoyed* and *ashamed* (A−),
  - through **marginally acceptable** *familiar* (B−) and *account* (C+),
  - to the **least acceptable** *speak* (C(−)), *suffer* (C(−)), and *talk* (D(+));
- **PP vs. CP difference** (according to linear mixed effects models):
  - not statistically significant for *annoyed* and *ashamed*,
  - statistically significant ($p < .001$) for the other six predicates;
- all **means of CP conditions** – with the exception of *talk* – **statistically significantly higher** than that of standard items D.

Length and complexity were controlled for, so it is **unlikely that these differences reflect processing**.

This justifies positing the following **hypothesis**:
Syntactic selectional restrictions may be gradient. **Corollary**: **Grammaticality is gradient.**

Experimental Evidence 3

Summary of experimental evidence:

- the acceptability of **PP variants was always high** (between A− and B+),
- the acceptability of **CP variants varied greatly**:
  - from **fully acceptable** *annoyed* and *ashamed* (A−),
  - through **marginally acceptable** *familiar* (B−) and *account* (C+),
  - to the **least acceptable** *speak* (C(−)), *suffer* (C(−)), and *talk* (D(+));
- **PP vs. CP difference** (according to linear mixed effects models):
  - not statistically significant for *annoyed* and *ashamed*,
  - statistically significant ($p < .001$) for the other six predicates;
- all **means of CP conditions** – with the exception of *talk* – **statistically significantly higher** than that of standard items D.

Length and complexity were controlled for, so it is **unlikely that these differences reflect processing**.

This justifies positing the following **hypothesis**:
Syntactic selectional restrictions may be gradient. **Corollary**: **Grammaticality is gradient.**

Experimental Evidence 3

Summary of experimental evidence:

- the acceptability of **PP variants was always high** (between A− and B+),
- the acceptability of **CP variants varied greatly**:
  - from **fully acceptable** *annoyed* and *ashamed* (A−),
  - through **marginally acceptable** *familiar* (B−) and *account* (C+),
  - to the **least acceptable** *speak* (C(−)), *suffer* (C(−)), and *talk* (D(+));
- **PP vs. CP difference** (according to linear mixed effects models):
  - not statistically significant for *annoyed* and *ashamed*,
  - statistically significant ($p < .001$) for the other six predicates;
- all **means of CP conditions** – with the exception of *talk* – **statistically significantly higher** than that of standard items D.

Length and complexity were controlled for, so it is **unlikely that these differences reflect processing**.

This justifies positing the following **hypothesis**:
Syntactic selectional restrictions may be gradient. **Corollary**: **Grammaticality is gradient.**

Experimental Evidence 3

Summary of experimental evidence:

- the acceptability of **PP variants was always high** (between A− and B+),
- the acceptability of **CP variants varied greatly**:
  - from **fully acceptable** *annoyed* and *ashamed* (A−),
  - through **marginally acceptable** *familiar* (B−) and *account* (C+),
  - to the **least acceptable** *speak* (C(−)), *suffer* (C(−)), and *talk* (D(+));
- **PP vs. CP difference** (according to linear mixed effects models):
  - not statistically significant for *annoyed* and *ashamed*,
  - statistically significant ($p < .001$) for the other six predicates;
- all **means of CP conditions** – with the exception of *talk* – **statistically significantly higher** than that of standard items D.

Length and complexity were controlled for, so it is **unlikely that these differences reflect processing**.

This justifies positing the following **hypothesis**:
Syntactic selectional restrictions may be gradient. **Corollary**: **Grammaticality is gradient.**

Experimental Evidence 3

Summary of experimental evidence:

- the acceptability of **PP variants was always high** (between A− and B+),
- the acceptability of **CP variants varied greatly**:
  - from **fully acceptable** *annoyed* and *ashamed* (A−),
  - through **marginally acceptable** *familiar* (B−) and *account* (C+),
  - to the **least acceptable** *speak* (C(−)), *suffer* (C(−)), and *talk* (D(+));
- **PP vs. CP difference** (according to linear mixed effects models):
  - not statistically significant for *annoyed* and *ashamed*,
  - statistically significant ($p < .001$) for the other six predicates;
- all **means of CP conditions** – with the exception of *talk* – **statistically significantly higher** than that of standard items D.

Length and complexity were controlled for, so it is **unlikely that these differences reflect processing**.

This justifies positing the following **hypothesis**:
Syntactic selectional restrictions may be gradient. Corollary: Grammaticality is gradient.

Experimental Evidence 3

Summary of experimental evidence:

- the acceptability of **PP variants was always high** (between A− and B+),
- the acceptability of **CP variants varied greatly**:
  - from **fully acceptable** *annoyed* and *ashamed* (A−),
  - through **marginally acceptable** *familiar* (B−) and *account* (C+),
  - to the **least acceptable** *speak* (C(−)), *suffer* (C(−)), and *talk* (D(+));
- **PP vs. CP difference** (according to linear mixed effects models):
  - not statistically significant for *annoyed* and *ashamed*,
  - statistically significant ($p < .001$) for the other six predicates;
- all **means of CP conditions** – with the exception of *talk* – **statistically significantly higher** than that of standard items D.

Length and complexity were controlled for, so it is **unlikely that these differences reflect processing**.

This justifies positing the following **hypothesis**:
Syntactic selectional restrictions may be gradient. **Corollary**: **Grammaticality is gradient.**

Supporting Corpus Evidence 1

The number of occurrences of predicate plus *that* in English Web 2021 and the number of **true combinations of the predicate with its CP dependent among a 100-hit random sample**:

Predicate	Hits	Sampled	True Positives
annoyed	501	100	**93**
ashamed	4,551	100	**85**
familiar	316	100	**48**
account	7,975	100	3
depend	534	100	**15**
speak	17,636	100	**10**
suffer	11,603	100	1
talk	13,732	100	6

Supporting Corpus Evidence 1

The number of occurrences of predicate plus *that* in English Web 2021 and the number of **true combinations of the predicate with its CP dependent among a 100-hit random sample**:

Predicate	Hits	Sampled	True Positives
annoyed	501	100	93
ashamed	4,551	100	85
familiar	316	100	48
account	7,975	100	3
depend	534	100	15
speak	17,636	100	10
suffer	11,603	100	1
talk	13,732	100	6

Supporting Corpus Evidence 2

Some **examples** for ***depend***:

1. *No matter what management does in the States, we always have to* ***depend*** [$_{\text{CP}}$ *that they are building it to our specs, which isn't always the case*].
2. *They respect each other's role and* ***depend*** [$_{\text{CP}}$ *that each wolf in the pack will live up to their individual responsibility*].
3. *…and you may* ***depend*** [$_{\text{CP}}$ *that we will endeavor to be your guard*].

Some for ***talk***:

4. *Some of us old timers were out at Avis Collier's Store not long ago and were* ***talking*** [$_{\text{CP}}$ *that we hadn't been to a Serenading Party for about 40 years or better*].
5. *…they were* ***talking*** [$_{\text{CP}}$ *that they're going to be the next champion of the world*].

Supporting Corpus Evidence 2

Some **examples** for ***depend***:

1. *No matter what management does in the States, we always have to* ***depend*** [$_{CP}$ *that they are building it to our specs, which isn't always the case*].
2. *They respect each other's role and* ***depend*** [$_{CP}$ *that each wolf in the pack will live up to their individual responsibility*].
3. ... *and you may* ***depend*** [$_{CP}$ *that we will endeavor to be your guard*].

Some for ***talk***:

4. *Some of us old timers were out at Avis Collier's Store not long ago and were* ***talking*** [$_{CP}$ *that we hadn't been to a Serenading Party for about 40 years or better*].
5. ... *they were* ***talking*** [$_{CP}$ *that they're going to be the next champion of the world*].

Supporting Corpus Evidence 2

Some **examples** for ***depend***:

1. *No matter what management does in the States, we always have to* ***depend*** [$_{\text{CP}}$ *that they are building it to our specs, which isn't always the case*].
2. *They respect each other's role and* ***depend*** [$_{\text{CP}}$ *that each wolf in the pack will live up to their individual responsibility*].
3. *... and you may* ***depend*** [$_{\text{CP}}$ *that we will endeavor to be your guard*].

Some for ***talk***:

4. *Some of us old timers were out at Avis Collier's Store not long ago and were* ***talking*** [$_{\text{CP}}$ *that we hadn't been to a Serenading Party for about 40 years or better*].
5. *... they were* ***talking*** [$_{\text{CP}}$ *that they're going to be the next champion of the world*].

Supporting Corpus Evidence 2

Some **examples** for ***depend***:

1. *No matter what management does in the States, we always have to* ***depend*** [$_{\text{CP}}$ *that they are building it to our specs, which isn't always the case*].
2. *They respect each other's role and* ***depend*** [$_{\text{CP}}$ *that each wolf in the pack will live up to their individual responsibility*].
3. *...and you may* ***depend*** [$_{\text{CP}}$ *that we will endeavor to be your guard*].

Some for ***talk***:

4. *Some of us old timers were out at Avis Collier's Store not long ago and were* ***talking*** [$_{\text{CP}}$ *that we hadn't been to a Serenading Party for about 40 years or better*].
5. *...they were* ***talking*** [$_{\text{CP}}$ *that they're going to be the next champion of the world*].

Gradient grammaticality in HPSG 1

Formally, an **HPSG grammar** is a **set θ of closed formulae** ("constraints") δ_i (Richter 2004, 2024: 173).

A **gradient HPSG grammar** (cf. Şenşekerci 2024) is:

- a **set θ of $\langle \delta_i, w_i \rangle$ pairs** such that:
  - δ_i is a **closed formula** ("constraint"), and
  - $w_i \in [0, 1]$ is **its weight** – the cost of a single violation of this constraint.

Some **scenarios**:

violations	burden for the model
no violations	0
a single object violates a single constraint δ_j	w_j
two objects violate δ_j and another violates δ_k	$2 \cdot w_j + w_k$

Gradient grammaticality in HPSG 1

Formally, an **HPSG grammar** is a **set θ of closed formulae** (“constraints”) δ_i (Richter 2004, 2024: 173).

A **gradient HPSG grammar** (cf. Şenşekerci 2024) is:

- a **set θ of $\langle \delta_i, w_i \rangle$ pairs** such that:
  - δ_i is a **closed formula** (“constraint”), and
  - $w_i \in [0, 1]$ is **its weight** – the cost of a single violation of this constraint.

Some **scenarios**:

violations	burden for the model
no violations	0
a single object violates a single constraint δ_j	w_j
two objects violate δ_j and another violates δ_k	$2 \cdot w_j + w_k$

Gradient grammaticality in HPSG 1

Formally, an **HPSG grammar** is a **set θ of closed formulae** (“constraints”) δ_i (Richter 2004, 2024: 173).

A **gradient HPSG grammar** (cf. Şenşekerci 2024) is:

- a **set θ of $\langle \delta_i, w_i \rangle$ pairs** such that:
  - δ_i is a **closed formula** (“constraint”), and
  - $w_i \in [0, 1]$ is **its weight** – the cost of a single violation of this constraint.

Some **scenarios**:

violations	burden for the model
no violations	0
a single object violates a single constraint δ_j	w_j
two objects violate δ_j and another violates δ_k	$2 \cdot w_j + w_k$

Gradient grammaticality in HPSG 1

Formally, an **HPSG grammar** is a **set θ of closed formulae** ("constraints") δ_i (Richter 2004, 2024: 173).

A **gradient HPSG grammar** (cf. Şenşekerci 2024) is:

- a **set θ of $\langle \delta_i, w_i \rangle$ pairs** such that:
  - δ_i is a **closed formula** ("constraint"), and
  - $w_i \in [0, 1]$ is **its weight** – the cost of a single violation of this constraint.

Some **scenarios**:

violations	burden for the model
no violations	0
a single object violates a single constraint δ_j	w_j
two objects violate δ_j and another violates δ_k	$2 \cdot w_j + w_k$

Gradient grammaticality in HPSG 1

Formally, an **HPSG grammar** is a **set θ of closed formulae** ("constraints") δ_i (Richter 2004, 2024: 173).

A **gradient HPSG grammar** (cf. Şenşekerci 2024) is:

- a **set θ of $\langle \delta_i, w_i \rangle$ pairs** such that:
  - δ_i is a **closed formula** ("constraint"), and
  - $w_i \in [0, 1]$ is **its weight** – the cost of a single violation of this constraint.

Some scenarios:

violations	burden for the model
no violations	0
a single object violates a single constraint δ_j	w_j
two objects violate δ_j and another violates δ_k	$2 \cdot w_j + w_k$

Gradient grammaticality in HPSG 1

Formally, an **HPSG grammar** is a **set θ of closed formulae** ("constraints") δ_i (Richter 2004, 2024: 173).

A **gradient HPSG grammar** (cf. Şenşekerci 2024) is:

- a **set θ of $\langle \delta_i, w_i \rangle$ pairs** such that:
  - δ_i is a **closed formula** ("constraint"), and
  - $w_i \in [0, 1]$ is **its weight** – the cost of a single violation of this constraint.

Some **scenarios**:

violations	burden for the model
no violations	0
a single object violates a single constraint δ_j	w_j
two objects violate δ_j and another violates δ_k	$2 \cdot w_j + w_k$

Gradient grammaticality in HPSG 1

Formally, an **HPSG grammar** is a **set θ of closed formulae** ("constraints") δ_i (Richter 2004, 2024: 173).

A **gradient HPSG grammar** (cf. Şenşekerci 2024) is:

- a **set θ of $\langle \delta_i, w_i \rangle$ pairs** such that:
  - δ_i is a **closed formula** ("constraint"), and
  - $w_i \in [0, 1]$ is **its weight** – the cost of a single violation of this constraint.

Some **scenarios**:

violations	burden for the model
no violations	0
a single object violates a single constraint δ_j	w_j
two objects violate δ_j and another violates δ_k	$2 \cdot w_j + w_k$

Gradient grammaticality in HPSG 1

Formally, an **HPSG grammar** is a **set θ of closed formulae** ("constraints") δ_i (Richter 2004, 2024: 173).

A **gradient HPSG grammar** (cf. Şenşekerci 2024) is:

- a **set θ of $\langle \delta_i, w_i \rangle$ pairs** such that:
  - δ_i is a **closed formula** ("constraint"), and
  - $w_i \in [0, 1]$ is **its weight** – the cost of a single violation of this constraint.

Some **scenarios**:

violations	burden for the model
no violations	0
a single object violates a single constraint δ_j	w_j
two objects violate δ_j and another violates δ_k	$2 \cdot w_j + w_k$

Gradient grammaticality in HPSG 1

Formally, an **HPSG grammar** is a **set θ of closed formulae** ("constraints") δ_i (Richter 2004, 2024: 173).

A **gradient HPSG grammar** (cf. Şenşekerci 2024) is:

- a **set θ of $\langle \delta_i, w_i \rangle$ pairs** such that:
  - δ_i is a **closed formula** ("constraint"), and
  - $w_i \in [0, 1]$ is **its weight** – the cost of a single violation of this constraint.

Some **scenarios**:

violations	burden for the model
no violations	0
a single object violates a single constraint δ_j	w_j
two objects violate δ_j and another violates δ_k	$2 \cdot w_j + w_k$

Gradient grammaticality in HPSG 2

In formula, the **burden for a model** I:

$$\mathrm{B(I)} = \sum_{\langle \delta_i, w_i \rangle \in \theta} |\mathrm{Viol}(\delta_i)| \cdot w_i$$

where $\mathrm{Viol}(\delta_i)$ is the set of model objects violating δ_i.

This **does not take into account** the sometimes reported ***under*additive effect of multiple violations**, in particular, "**floor effects**".

A formula that does:

$$\mathrm{B(I)} = 1 - \prod_{\langle \delta_i, w_i \rangle \in \theta} (1 - w_i)^{|\mathrm{Viol}(\delta_i)|}$$

Examples:

- - model I_1 with **three weighty violations** worth 0.7, 0.8, and 0.9
  - $\mathrm{B(I_1)} = 1 - (1 - 0.7) \cdot (1 - 0.8) \cdot (1 - 0.9) = 0.994 \approx 1.0$
  - the "floor effect" in acceptability judgements
- - model I_2 with **a double violation of a very light constraint** worth 0.05
  - $\mathrm{B(I_2)} = 1 - (1 - 0.05)^2 = 0.0975 \approx 0.1$
  - an *almost fully* additive small burden

Gradient grammaticality in HPSG 2

In formula, the **burden for a model** I:

$$B(I) = \sum_{\langle \delta_i, w_i \rangle \in \theta} |Viol(\delta_i)| \cdot w_i$$

where $Viol(\delta_i)$ is the set of model objects violating δ_i.

This **does not take into account** the sometimes reported ***under*****additive effect of multiple violations**, in particular, "**floor effects**".

A formula that does:

$$B(I) = 1 - \prod_{\langle \delta_i, w_i \rangle \in \theta} (1 - w_i)^{|Viol(\delta_i)|}$$

Examples:

- model I_1 with **three weighty violations** worth 0.7, 0.8, and 0.9
  - $B(I_1) = 1 - (1 - 0.7) \cdot (1 - 0.8) \cdot (1 - 0.9) = 0.994 \approx 1.0$
  - the "floor effect" in acceptability judgements
- model I_2 with **a double violation of a very light constraint** worth 0.05
  - $B(I_2) = 1 - (1 - 0.05)^2 = 0.0975 \approx 0.1$
  - an *almost fully* additive small burden

Gradient grammaticality in HPSG 2

In formula, the **burden for a model** I:

$$\mathrm{B}(\mathrm{I}) = \sum_{\langle \delta_i, w_i \rangle \in \theta} |\mathrm{Viol}(\delta_i)| \cdot w_i$$

where $\mathrm{Viol}(\delta_i)$ is the set of model objects violating δ_i.

This **does not take into account** the sometimes reported ***under*additive effect of multiple violations**, in particular, "**floor effects**".

A formula that does:

$$\mathrm{B}(\mathrm{I}) = 1 - \prod_{\langle \delta_i, w_i \rangle \in \theta} (1 - w_i)^{|\mathrm{Viol}(\delta_i)|}$$

Examples:

- model I_1 with **three weighty violations** worth 0.7, 0.8, and 0.9
  - $\mathrm{B}(\mathrm{I}_1) = 1 - (1 - 0.7) \cdot (1 - 0.8) \cdot (1 - 0.9) = 0.994 \approx 1.0$
  - the "floor effect" in acceptability judgements
- model I_2 with **a double violation of a very light constraint** worth 0.05
  - $\mathrm{B}(\mathrm{I}_2) = 1 - (1 - 0.05)^2 = 0.0975 \approx 0.1$
  - an *almost fully* additive small burden

Gradient grammaticality in HPSG 2

In formula, the **burden for a model** I:

$$\mathrm{B}(\mathrm{I}) = \sum_{\langle \delta_i, w_i \rangle \in \theta} |\mathrm{Viol}(\delta_i)| \cdot w_i$$

where $\mathrm{Viol}(\delta_i)$ is the set of model objects violating δ_i.

This **does not take into account** the sometimes reported ***under*****additive effect of multiple violations**, in particular, "**floor effects**".

A formula that does:

$$\mathrm{B}(\mathrm{I}) = 1 - \prod_{\langle \delta_i, w_i \rangle \in \theta} (1 - w_i)^{|\mathrm{Viol}(\delta_i)|}$$

Examples:

- model I_1 with **three weighty violations** worth 0.7, 0.8, and 0.9
  - $\mathrm{B}(\mathrm{I}_1) = 1 - (1 - 0.7) \cdot (1 - 0.8) \cdot (1 - 0.9) = 0.994 \approx 1.0$
  - the "floor effect" in acceptability judgements
- model I_2 with **a double violation of a very light constraint** worth 0.05
  - $\mathrm{B}(\mathrm{I}_2) = 1 - (1 - 0.05)^2 = 0.0975 \approx 0.1$
  - an *almost fully* additive small burden

Gradient grammaticality in HPSG 2

In formula, the **burden for a model** I:

$$\mathsf{B}(\mathsf{I}) = \sum_{\langle \delta_i, w_i \rangle \in \theta} |\mathsf{Viol}(\delta_i)| \cdot w_i$$

where $\mathsf{Viol}(\delta_i)$ is the set of model objects violating δ_i.

This **does not take into account** the sometimes reported ***under*additive effect of multiple violations**, in particular, "**floor effects**".

A formula that does:

$$\mathsf{B}(\mathsf{I}) = 1 - \prod_{\langle \delta_i, w_i \rangle \in \theta} (1 - w_i)^{|\mathsf{Viol}(\delta_i)|}$$

Examples:

- model I_1 with **three weighty violations** worth 0.7, 0.8, and 0.9
  - $\mathsf{B}(\mathsf{I}_1) = 1 - (1 - 0.7) \cdot (1 - 0.8) \cdot (1 - 0.9) = 0.994 \approx 1.0$
  - the "floor effect" in acceptability judgements
- model I_2 with **a double violation of a very light constraint** worth 0.05
  - $\mathsf{B}(\mathsf{I}_2) = 1 - (1 - 0.05)^2 = 0.0975 \approx 0.1$
  - an *almost fully* additive small burden

Gradient grammaticality in HPSG 2

In formula, the **burden for a model** I:

$$B(I) = \sum_{\langle \delta_i, w_i \rangle \in \theta} |Viol(\delta_i)| \cdot w_i$$

where $Viol(\delta_i)$ is the set of model objects violating δ_i.

This **does not take into account** the sometimes reported ***under*****additive effect of multiple violations**, in particular, "**floor effects**".

A formula that does:

$$B(I) = 1 - \prod_{\langle \delta_i, w_i \rangle \in \theta} (1 - w_i)^{|Viol(\delta_i)|}$$

Examples:

- model I_1 with **three weighty violations** worth 0.7, 0.8, and 0.9
  - $B(I_1) = 1 - (1 - 0.7) \cdot (1 - 0.8) \cdot (1 - 0.9) = 0.994 \approx 1.0$
  - the "floor effect" in acceptability judgements
- model I_2 with **a double violation of a very light constraint** worth 0.05
  - $B(I_2) = 1 - (1 - 0.05)^2 = 0.0975 \approx 0.1$
  - an *almost fully* additive small burden

Gradient grammaticality in HPSG 2

In formula, the **burden for a model** I:

$$\mathrm{B(I)} = \sum_{\langle \delta_i, w_i \rangle \in \theta} |\mathrm{Viol}(\delta_i)| \cdot w_i$$

where $\mathrm{Viol}(\delta_i)$ is the set of model objects violating δ_i.

This **does not take into account** the sometimes reported ***under*additive effect of multiple violations**, in particular, "**floor effects**".

A formula that does:

$$\mathrm{B(I)} = 1 - \prod_{\langle \delta_i, w_i \rangle \in \theta} (1 - w_i)^{|\mathrm{Viol}(\delta_i)|}$$

Examples:

- - model I_1 with **three weighty violations** worth 0.7, 0.8, and 0.9
  - $\mathrm{B(I_1)} = 1 - (1 - 0.7) \cdot (1 - 0.8) \cdot (1 - 0.9) = 0.994 \approx 1.0$
  - the "floor effect" in acceptability judgements
- - model I_2 with **a double violation of a very light constraint** worth 0.05
  - $\mathrm{B(I_2)} = 1 - (1 - 0.05)^2 = 0.0975 \approx 0.1$
  - an *almost fully* additive small burden

Gradient grammaticality in HPSG 2

In formula, the **burden for a model** I:

$$\mathsf{B}(\mathsf{I}) = \sum_{\langle \delta_i, w_i \rangle \in \theta} |\mathsf{Viol}(\delta_i)| \cdot w_i$$

where $\mathsf{Viol}(\delta_i)$ is the set of model objects violating δ_i.

This **does not take into account** the sometimes reported ***under*****additive effect of multiple violations**, in particular, "**floor effects**".

A formula that does:

$$\mathsf{B}(\mathsf{I}) = 1 - \prod_{\langle \delta_i, w_i \rangle \in \theta} (1 - w_i)^{|\mathsf{Viol}(\delta_i)|}$$

Examples:

- model I_1 with **three weighty violations** worth 0.7, 0.8, and 0.9
  - $\mathsf{B}(\mathsf{I}_1) = 1 - (1 - 0.7) \cdot (1 - 0.8) \cdot (1 - 0.9) = 0.994 \approx 1.0$
  - the "floor effect" in acceptability judgements
- model I_2 with **a double violation of a very light constraint** worth 0.05
  - $\mathsf{B}(\mathsf{I}_2) = 1 - (1 - 0.05)^2 = 0.0975 \approx 0.1$
  - an *almost fully* additive small burden

Gradient grammaticality in HPSG 2

In formula, the **burden for a model** I:

$$B(I) = \sum_{\langle \delta_i, w_i \rangle \in \theta} |\mathrm{Viol}(\delta_i)| \cdot w_i$$

where $\mathrm{Viol}(\delta_i)$ is the set of model objects violating δ_i.

This **does not take into account** the sometimes reported ***under*additive effect of multiple violations**, in particular, "**floor effects**".

A formula that does:

$$B(I) = 1 - \prod_{\langle \delta_i, w_i \rangle \in \theta} (1 - w_i)^{|\mathrm{Viol}(\delta_i)|}$$

Examples:

- - model I_1 with **three weighty violations** worth 0.7, 0.8, and 0.9
  - $B(I_1) = 1 - (1 - 0.7) \cdot (1 - 0.8) \cdot (1 - 0.9) = 0.994 \approx 1.0$
  - the "floor effect" in acceptability judgements
- - model I_2 with **a double violation of a very light constraint** worth 0.05
  - $B(I_2) = 1 - (1 - 0.05)^2 = 0.0975 \approx 0.1$
  - an *almost fully* additive small burden

Gradient grammaticality in HPSG 2

In formula, the **burden for a model** I:

$$\mathrm{B}(\mathrm{I}) = \sum_{\langle \delta_i, w_i \rangle \in \theta} |\mathrm{Viol}(\delta_i)| \cdot w_i$$

where $\mathrm{Viol}(\delta_i)$ is the set of model objects violating δ_i.

This **does not take into account** the sometimes reported ***under*additive effect of multiple violations**, in particular, "**floor effects**".

A formula that does:

$$\mathrm{B}(\mathrm{I}) = 1 - \prod_{\langle \delta_i, w_i \rangle \in \theta} (1 - w_i)^{|\mathrm{Viol}(\delta_i)|}$$

Examples:

- - model I_1 with **three weighty violations** worth 0.7, 0.8, and 0.9
  - $\mathrm{B}(\mathrm{I}_1) = 1 - (1 - 0.7) \cdot (1 - 0.8) \cdot (1 - 0.9) = 0.994 \approx 1.0$
  - the "floor effect" in acceptability judgements
- - model I_2 with **a double violation of a very light constraint** worth 0.05
  - $\mathrm{B}(\mathrm{I}_2) = 1 - (1 - 0.05)^2 = 0.0975 \approx 0.1$
  - an *almost fully* additive small burden

Gradient grammaticality in HPSG 2

In formula, the **burden for a model** I:

$$\mathrm{B(I)} = \sum_{\langle \delta_i, w_i \rangle \in \theta} |\mathrm{Viol}(\delta_i)| \cdot w_i$$

where $\mathrm{Viol}(\delta_i)$ is the set of model objects violating δ_i.

This **does not take into account** the sometimes reported ***under*****additive effect of multiple violations**, in particular, "**floor effects**".

A formula that does:

$$\mathrm{B(I)} = 1 - \prod_{\langle \delta_i, w_i \rangle \in \theta} (1 - w_i)^{|\mathrm{Viol}(\delta_i)|}$$

Examples:

- - model I_1 with **three weighty violations** worth 0.7, 0.8, and 0.9
  - $\mathrm{B(I_1)} = 1 - (1 - 0.7) \cdot (1 - 0.8) \cdot (1 - 0.9) = 0.994 \approx 1.0$
  - the "floor effect" in acceptability judgements
- - model I_2 with **a double violation of a very light constraint** worth 0.05
  - $\mathrm{B(I_2)} = 1 - (1 - 0.05)^2 = 0.0975 \approx 0.1$
  - an *almost fully* additive small burden

Selectional restrictions in HPSG

Assume for simplicity that idiosyncratic **syntactic selectional restrictions** are directly **encoded in lexical entries**, e.g.:

- $\mathrm{LE}_{depend} \stackrel{df}{\equiv} \begin{bmatrix} \textsc{phon} \ \langle \textit{depend} \rangle \\ \ldots \\ \textsc{arg-st} \ \langle \mathrm{NP}[\textsc{nom}], \mathrm{PP}[\textit{on}] \rangle \end{bmatrix}$

Such lexical entries are built into **a larger "lexicon" constraint**, e.g., the **Word Principle** (Höhle 1994, Richter 2024: 101):

- $\textit{word} \rightarrow (\mathrm{LE}_{depend} \vee \mathrm{LE}_{Marge} \vee \mathrm{LE}_{wait} \vee \ldots)$

Problem:
How to say that the 2nd argument of *depend* may be a CP[*that*], but with some penalty?

Non-starter:
A weight attached to the Word Principle above only reflects the general penalty for using an out-of-vocabulary word.

Selectional restrictions in HPSG

Assume for simplicity that idiosyncratic **syntactic selectional restrictions** are directly **encoded in lexical entries**, e.g.:

- $\mathrm{LE}_{\mathit{depend}} \overset{df}{\equiv} \begin{bmatrix} \textsc{phon} & \langle \mathit{depend} \rangle \\ \ldots & \\ \textsc{arg-st} & \langle \mathrm{NP}[\textsc{nom}], \mathrm{PP}[\mathit{on}] \rangle \end{bmatrix}$

Such lexical entries are built into **a larger "lexicon" constraint**, e.g., the **Word Principle** (Höhle 1994, Richter 2024: 101):

- $\mathit{word} \rightarrow (\mathrm{LE}_{\mathit{depend}} \vee \mathrm{LE}_{\mathit{Marge}} \vee \mathrm{LE}_{\mathit{wait}} \vee \ldots)$

Problem:
How to say that the 2nd argument of *depend* may be a CP[*that*], but with some penalty?

Non-starter:
A weight attached to the Word Principle above only reflects the general penalty for using an out-of-vocabulary word.

Selectional restrictions in HPSG

Assume for simplicity that idiosyncratic **syntactic selectional restrictions** are directly **encoded in lexical entries**, e.g.:

- $\mathrm{LE}_{depend} \stackrel{df}{\equiv} \begin{bmatrix} \textsc{phon} \ \langle \textit{depend} \rangle \\ \ldots \\ \textsc{arg-st} \ \langle \mathrm{NP}[\textsc{nom}], \mathrm{PP}[\textit{on}] \rangle \end{bmatrix}$

Such lexical entries are built into **a larger "lexicon" constraint**, e.g., the Word Principle (Höhle 1994, Richter 2024: 101):

- $\textit{word} \rightarrow (\mathrm{LE}_{depend} \vee \mathrm{LE}_{Marge} \vee \mathrm{LE}_{wait} \vee \ldots)$

Problem:
How to say that the 2nd argument of *depend* may be a CP[*that*], but with some penalty?

Non-starter:
A weight attached to the Word Principle above only reflects the general penalty for using an out-of-vocabulary word.

Selectional restrictions in HPSG

Assume for simplicity that idiosyncratic **syntactic selectional restrictions** are directly **encoded in lexical entries**, e.g.:

- $\mathrm{LE}_{\textit{depend}} \overset{df}{\equiv} \begin{bmatrix} \textsc{phon} \ \langle \textit{depend} \rangle \\ \ldots \\ \textsc{arg-st} \ \langle \mathrm{NP}[\textsc{nom}], \mathrm{PP}[\textit{on}] \rangle \end{bmatrix}$

Such lexical entries are built into **a larger "lexicon" constraint**, e.g., the **Word Principle** (Höhle 1994, Richter 2024: 101):

- $\textit{word} \rightarrow (\mathrm{LE}_{\textit{depend}} \vee \mathrm{LE}_{\textit{Marge}} \vee \mathrm{LE}_{\textit{wait}} \vee \ldots)$

Problem:
How to say that the 2nd argument of *depend* may be a CP[*that*], but with some penalty?

Non-starter:
A weight attached to the Word Principle above only reflects the general penalty for using an out-of-vocabulary word.

Selectional restrictions in HPSG

Assume for simplicity that idiosyncratic **syntactic selectional restrictions** are directly **encoded in lexical entries**, e.g.:

- $\mathrm{LE}_{\mathit{depend}} \overset{df}{\equiv} \begin{bmatrix} \textsc{phon} \ \langle \mathit{depend} \rangle \\ \ldots \\ \textsc{arg-st} \ \langle \mathrm{NP}[\textsc{nom}], \mathrm{PP}[\mathit{on}] \rangle \end{bmatrix}$

Such lexical entries are built into **a larger "lexicon" constraint**, e.g., the **Word Principle** (Höhle 1994, Richter 2024: 101):

- $\mathit{word} \rightarrow (\mathrm{LE}_{\mathit{depend}} \vee \mathrm{LE}_{\mathit{Marge}} \vee \mathrm{LE}_{\mathit{wait}} \vee \ldots)$

Problem:
How to say that the 2nd argument of *depend* may be a CP[*that*], but with some penalty?

Non-starter:
A weight attached to the Word Principle above only reflects the general penalty for using an out-of-vocabulary word.

Selectional restrictions in HPSG

Assume for simplicity that idiosyncratic **syntactic selectional restrictions** are directly **encoded in lexical entries**, e.g.:

- $\mathrm{LE}_{\mathit{depend}} \stackrel{df}{\equiv} \begin{bmatrix} \textsc{phon} & \langle \mathit{depend} \rangle \\ \ldots & \\ \textsc{arg-st} & \langle \mathrm{NP}[\textsc{nom}], \mathrm{PP}[\mathit{on}] \rangle \end{bmatrix}$

Such lexical entries are built into **a larger "lexicon" constraint**, e.g., the **Word Principle** (Höhle 1994, Richter 2024: 101):

- $\mathit{word} \rightarrow (\mathrm{LE}_{\mathit{depend}} \vee \mathrm{LE}_{\mathit{Marge}} \vee \mathrm{LE}_{\mathit{wait}} \vee \ldots)$

Problem:
How to say that the 2nd argument of *depend* may be a CP[*that*], but with some penalty?

Non-starter:
A weight attached to the Word Principle above only reflects the general penalty for using an out-of-vocabulary word.

Gradient selectional restrictions in HPSG 1

Two solutions:

1. one based on the assumption that there is a **dichotomy between hard and soft selectional restrictions**,
2. another **not making such an assumption**.

Solution 1:

- **lexical entries mention all possibilities**, also those violating soft constraints, e.g.:
- $\mathrm{LE}_{depend} \stackrel{df}{\equiv} \begin{bmatrix} \textsc{phon}\ \langle depend \rangle \\ \ldots \\ \textsc{arg-st}\ \langle \mathrm{NP}[\textsc{nom}],\ \mathrm{PP}[on] \vee \mathrm{CP}[that] \rangle \end{bmatrix}$
- note the lack of an NP in the 2nd argument (a hard constraint), e.g.:
- **Justin Bieber can* ***depend*** [$_{\mathrm{NP}}$ *the fact* [$_{\mathrm{CP}}$ *that his fans still love his early songs*]].
- then add **separate soft constraints** on particular realizations of particular arguments, e.g.:
- $\left\langle \begin{bmatrix} \textsc{phon}\ \langle depend \rangle \\ \textsc{arg-st}\ \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix} \longrightarrow \neg\boxed{2}\left[\ldots|\textsc{head}\ \ complementiser\right],\ \ 0.28 \right\rangle$ ($\boxed{2}$ is not a CP)

Gradient selectional restrictions in HPSG 1

Two solutions:

1. one based on the assumption that there is a **dichotomy between hard and soft selectional restrictions**,
2. another **not making such an assumption**.

Solution 1:

- **lexical entries mention all possibilities**, also those violating soft constraints, e.g.:
- $\mathrm{LE}_{depend} \stackrel{df}{\equiv} \begin{bmatrix} \textsc{phon}\ \langle \textit{depend} \rangle \\ \dots \\ \textsc{arg-st}\ \langle \mathrm{NP}[\textsc{nom}],\ \mathrm{PP}[\textit{on}] \vee \mathrm{CP}[\textit{that}] \rangle \end{bmatrix}$
- note the lack of an NP in the 2nd argument (a hard constraint), e.g.:
- **Justin Bieber can* ***depend*** [$_{\mathrm{NP}}$ *the fact* [$_{\mathrm{CP}}$ *that his fans still love his early songs*]].
- then add **separate soft constraints** on particular realizations of particular arguments, e.g.:
- $\left\langle \begin{bmatrix} \textsc{phon}\ \langle \textit{depend} \rangle \\ \textsc{arg-st}\ \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix} \longrightarrow \neg \boxed{2} [\dots | \textsc{head}\ \textit{complementiser}],\ 0.28 \right\rangle$ ($\boxed{2}$ is not a CP)

Gradient selectional restrictions in HPSG 1

Two solutions:

1. one based on the assumption that there is a **dichotomy between hard and soft selectional restrictions**,
2. another **not making such an assumption**.

Solution 1:

- **lexical entries mention all possibilities**, also those violating soft constraints, e.g.:
- $\mathrm{LE}_{depend} \stackrel{df}{\equiv} \begin{bmatrix} \text{PHON } \langle \textit{depend} \rangle \\ \dots \\ \text{ARG-ST } \langle \text{NP[NOM]}, \text{PP}[\textit{on}] \vee \text{CP}[\textit{that}] \rangle \end{bmatrix}$
- note the lack of an NP in the 2nd argument (a hard constraint), e.g.:
- **Justin Bieber can* ***depend*** [$_{\text{NP}}$ *the fact* [$_{\text{CP}}$ *that his fans still love his early songs*]].
- then add **separate soft constraints** on particular realizations of particular arguments, e.g.:
- $\left\langle \begin{bmatrix} \text{PHON } \langle \textit{depend} \rangle \\ \text{ARG-ST } \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix} \rightarrow \neg \boxed{2} [\dots | \text{HEAD } \textit{complementiser}],\ 0.28 \right\rangle$ ($\boxed{2}$ is not a CP)

Gradient selectional restrictions in HPSG 1

Two solutions:

1. one based on the assumption that there is a **dichotomy between hard and soft selectional restrictions**,
2. another **not making such an assumption**.

Solution 1:

- **lexical entries mention all possibilities**, also those violating soft constraints, e.g.:
- $\mathrm{LE}_{\textit{depend}} \stackrel{df}{\equiv} \begin{bmatrix} \text{PHON } \langle \textit{depend} \rangle \\ \ldots \\ \text{ARG-ST } \langle \mathrm{NP}[\textsc{nom}], \mathrm{PP}[\textit{on}] \vee \mathrm{CP}[\textit{that}] \rangle \end{bmatrix}$
- note the lack of an NP in the 2nd argument (a hard constraint), e.g.:
- **Justin Bieber can **depend** [$_{NP}$ the fact [$_{CP}$ that his fans still love his early songs]].*
- then add **separate soft constraints** on particular realizations of particular arguments, e.g.:
- $\left\langle \begin{bmatrix} \text{PHON } \langle \textit{depend} \rangle \\ \text{ARG-ST } \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix} \rightarrow \neg \boxed{2} [\ldots | \text{HEAD } \textit{complementiser}],\ 0.28 \right\rangle$ ($\boxed{2}$ is not a CP)

Gradient selectional restrictions in HPSG 1

Two solutions:

1. one based on the assumption that there is a **dichotomy between hard and soft selectional restrictions**,
2. another **not making such an assumption**.

Solution 1:

- **lexical entries mention all possibilities**, also those violating soft constraints, e.g.:
- $\text{LE}_{\textit{depend}} \stackrel{df}{\equiv} \begin{bmatrix} \text{PHON} \ \langle \textit{depend} \rangle \\ \ldots \\ \text{ARG-ST} \ \langle \text{NP[NOM]}, \text{PP}[\textit{on}] \vee \text{CP}[\textit{that}] \rangle \end{bmatrix}$
- note the lack of an NP in the 2nd argument (a hard constraint), e.g.:
- **Justin Bieber can* ***depend*** [$_{\text{NP}}$ *the fact* [$_{\text{CP}}$ *that his fans still love his early songs*]].
- then add **separate soft constraints** on particular realizations of particular arguments, e.g.:
- $\left\langle \begin{bmatrix} \text{PHON} \ \langle \textit{depend} \rangle \\ \text{ARG-ST} \ \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix} \rightarrow \neg \boxed{2} \left[\ldots | \text{HEAD} \ \textit{complementiser} \right], \ 0.28 \right\rangle$ ($\boxed{2}$ is not a CP)

Gradient selectional restrictions in HPSG 1

Two solutions:

1. one based on the assumption that there is a **dichotomy between hard and soft selectional restrictions**,
2. another **not making such an assumption**.

Solution 1:

- **lexical entries mention all possibilities**, also those violating soft constraints, e.g.:
- $\mathrm{LE}_{depend} \overset{df}{\equiv} \begin{bmatrix} \text{PHON} \ \langle \textit{depend} \rangle \\ \dots \\ \text{ARG-ST} \ \langle \text{NP[NOM]}, \text{PP}[\textit{on}] \vee \text{CP}[\textit{that}] \rangle \end{bmatrix}$
- note the lack of an NP in the 2nd argument (a hard constraint), e.g.:
- **Justin Bieber can* ***depend*** [$_{\text{NP}}$ *the fact* [$_{\text{CP}}$ *that his fans still love his early songs*]].
- then add **separate soft constraints** on particular realizations of particular arguments, e.g.:
- $\left\langle \begin{bmatrix} \text{PHON} \ \langle \textit{depend} \rangle \\ \text{ARG-ST} \ \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix} \rightarrow \neg\boxed{2}\left[\dots|\text{HEAD} \ \textit{complementiser}\right], \ 0.28 \right\rangle$ ($\boxed{2}$ is not a CP)

Gradient selectional restrictions in HPSG 1

Two solutions:

1. one based on the assumption that there is a **dichotomy between hard and soft selectional restrictions**,
2. another **not making such an assumption**.

Solution 1:

- **lexical entries mention all possibilities**, also those violating soft constraints, e.g.:
- $\text{LE}_{\textit{depend}} \overset{df}{\equiv} \begin{bmatrix} \textsc{phon} \ \langle \textit{depend} \rangle \\ \ldots \\ \textsc{arg-st} \ \langle \text{NP}[\textsc{nom}], \text{PP}[\textit{on}] \vee \text{CP}[\textit{that}] \rangle \end{bmatrix}$
- note the lack of an NP in the 2nd argument (a hard constraint), e.g.:
- **Justin Bieber can* ***depend*** [$_{\text{NP}}$ *the fact* [$_{\text{CP}}$ *that his fans still love his early songs*]].
- then add **separate soft constraints** on particular realizations of particular arguments, e.g.:
- $\left\langle \begin{bmatrix} \textsc{phon} \ \langle \textit{depend} \rangle \\ \textsc{arg-st} \ \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix} \longrightarrow \neg \boxed{2} [\ldots | \textsc{head} \ \textit{complementiser}], \ 0.28 \right\rangle$ ($\boxed{2}$ is not a CP)

Gradient selectional restrictions in HPSG 1

Two solutions:

1. one based on the assumption that there is a **dichotomy between hard and soft selectional restrictions**,
2. another **not making such an assumption**.

Solution 1:

- **lexical entries mention all possibilities**, also those violating soft constraints, e.g.:
- $\text{LE}_{\textit{depend}} \stackrel{df}{\equiv} \begin{bmatrix} \text{PHON} & \langle \textit{depend} \rangle \\ \ldots & \\ \text{ARG-ST} & \langle \text{NP[NOM]}, \text{PP}[\textit{on}] \vee \text{CP}[\textit{that}] \rangle \end{bmatrix}$
- note the lack of an NP in the 2nd argument (a hard constraint), e.g.:
- **Justin Bieber can* ***depend*** [$_{\text{NP}}$ *the fact* [$_{\text{CP}}$ *that his fans still love his early songs*]].
- then add **separate soft constraints** on particular realizations of particular arguments, e.g.:
- $\left\langle \begin{bmatrix} \text{PHON} & \langle \textit{depend} \rangle \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix} \rightarrow \neg \boxed{2} [\ldots | \text{HEAD} \ \textit{complementiser}], \ 0.28 \right\rangle$ ($\boxed{2}$ is not a CP)

Gradient selectional restrictions in HPSG 1

Two solutions:

1. one based on the assumption that there is a **dichotomy between hard and soft selectional restrictions**,
2. another **not making such an assumption**.

Solution 1:

- **lexical entries mention all possibilities**, also those violating soft constraints, e.g.:
- $\mathrm{LE}_{depend} \stackrel{df}{\equiv} \begin{bmatrix} \textsc{phon} \ \langle \textit{depend} \rangle \\ \ldots \\ \textsc{arg-st} \ \langle \mathrm{NP}[\textsc{nom}], \mathrm{PP}[\textit{on}] \vee \mathrm{CP}[\textit{that}] \rangle \end{bmatrix}$
- note the lack of an NP in the 2nd argument (a hard constraint), e.g.:
- **Justin Bieber can **depend** [$_{\mathrm{NP}}$ the fact [$_{\mathrm{CP}}$ that his fans still love his early songs]].*
- then add **separate soft constraints** on particular realizations of particular arguments, e.g.:
- $\left\langle \begin{bmatrix} \textsc{phon} \ \langle \textit{depend} \rangle \\ \textsc{arg-st} \ \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix} \longrightarrow \neg \boxed{2} \left[\ldots \middle| \textsc{head} \ \textit{complementiser} \right], \ 0.28 \right\rangle$ ($\boxed{2}$ is not a CP)

Gradient selectional restrictions in HPSG 1

Two solutions:

1. one based on the assumption that there is a **dichotomy between hard and soft selectional restrictions**,
2. another **not making such an assumption**.

Solution 1:

- **lexical entries mention all possibilities**, also those violating soft constraints, e.g.:
- $\text{LE}_{\textit{depend}} \overset{df}{\equiv} \begin{bmatrix} \text{PHON} \ \langle \textit{depend} \rangle \\ \ldots \\ \text{ARG-ST} \ \langle \text{NP}[\text{NOM}], \text{PP}[\textit{on}] \vee \text{CP}[\textit{that}] \rangle \end{bmatrix}$
- note the lack of an NP in the 2nd argument (a hard constraint), e.g.:
- **Justin Bieber can* ***depend*** [$_{\text{NP}}$ *the fact* [$_{\text{CP}}$ *that his fans still love his early songs*]].
- then add **separate soft constraints** on particular realizations of particular arguments, e.g.:
- $\left\langle \begin{bmatrix} \text{PHON} \ \langle \textit{depend} \rangle \\ \text{ARG-ST} \ \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix} \rightarrow \neg \boxed{2} \left[\ldots \middle| \text{HEAD} \ \textit{complementiser} \right], \ 0.28 \right\rangle$ ($\boxed{2}$ is not a CP)

Gradient selectional restrictions in HPSG 2

Solution 2 (more general, preferred):

- **lexical entries do not mention syntactic selectional restrictions** at all, only the number of arguments – or not even this, given some linking theory (Davis *et al.* 2024), e.g.:
- $\mathrm{LE}_{\textit{depend}} \stackrel{df}{\equiv} \begin{bmatrix} \textsc{phon} & \langle \textit{depend} \rangle \\ \ldots & \\ \textsc{arg-st} & \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix}$
- then add **separate constraints** on particular realizations of particular arguments, e.g.:
  - $\left\langle \begin{bmatrix} \textsc{phon} & \langle \textit{depend} \rangle \\ \textsc{arg-st} & \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix} \rightarrow \neg \boxed{2} \big[\ldots | \textsc{head}\ \textit{preposition} \big],\ 0.00 \right\rangle$ ($\boxed{2}$ is not a PP)
  - $\left\langle \begin{bmatrix} \textsc{phon} & \langle \textit{depend} \rangle \\ \textsc{arg-st} & \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix} \rightarrow \neg \boxed{2} \big[\ldots | \textsc{head}\ \textit{complementiser} \big],\ 0.28 \right\rangle$ ($\boxed{2}$ is not a CP)
  - $\left\langle \begin{bmatrix} \textsc{phon} & \langle \textit{depend} \rangle \\ \textsc{arg-st} & \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix} \rightarrow \neg \boxed{2} \big[\ldots | \textsc{head}\ \textit{noun} \big],\ 0.93 \right\rangle$ ($\boxed{2}$ is not an NP)
- such constraints could be stated at the level of **word classes** rather than particular words (cf. Davis *et al.* 2024: 351), perhaps within more general linking rules.

Gradient selectional restrictions in HPSG 2

Solution 2 (more general, preferred):

- **lexical entries do not mention syntactic selectional restrictions** at all, only the number of arguments – or not even this, given some linking theory (Davis *et al.* 2024), e.g.:
- $\mathrm{LE}_{\mathit{depend}} \overset{df}{\equiv} \begin{bmatrix} \textsc{phon} & \langle \mathit{depend} \rangle \\ \ldots & \\ \textsc{arg-st} & \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix}$
- then add **separate constraints** on particular realizations of particular arguments, e.g.:
  - $\left\langle \begin{bmatrix} \textsc{phon} & \langle \mathit{depend} \rangle \\ \textsc{arg-st} & \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix} \rightarrow \neg\boxed{2}\left[\ldots|\textsc{head}\ \mathit{preposition}\right],\ 0.00 \right\rangle$ ($\boxed{2}$ is not a PP)
  - $\left\langle \begin{bmatrix} \textsc{phon} & \langle \mathit{depend} \rangle \\ \textsc{arg-st} & \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix} \rightarrow \neg\boxed{2}\left[\ldots|\textsc{head}\ \mathit{complementiser}\right],\ 0.28 \right\rangle$ ($\boxed{2}$ is not a CP)
  - $\left\langle \begin{bmatrix} \textsc{phon} & \langle \mathit{depend} \rangle \\ \textsc{arg-st} & \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix} \rightarrow \neg\boxed{2}\left[\ldots|\textsc{head}\ \mathit{noun}\right],\ 0.93 \right\rangle$ ($\boxed{2}$ is not an NP)
- such constraints could be stated at the level of **word classes** rather than particular words (cf. Davis *et al.* 2024: 351), perhaps within more general linking rules.

Gradient selectional restrictions in HPSG 2

Solution 2 (more general, preferred):

- **lexical entries do not mention syntactic selectional restrictions** at all, only the number of arguments – or not even this, given some linking theory (Davis *et al.* 2024), e.g.:
- $\text{LE}_{depend} \stackrel{df}{\equiv} \begin{bmatrix} \text{PHON} & \langle \textit{depend} \rangle \\ \ldots & \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix}$
- then add **separate constraints** on particular realizations of particular arguments, e.g.:
  - $\left\langle \begin{bmatrix} \text{PHON} & \langle \textit{depend} \rangle \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix} \rightarrow \neg \boxed{2}[\ldots|\text{HEAD } \textit{preposition}],\ 0.00 \right\rangle$ ($\boxed{2}$ is not a PP)
  - $\left\langle \begin{bmatrix} \text{PHON} & \langle \textit{depend} \rangle \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix} \rightarrow \neg \boxed{2}[\ldots|\text{HEAD } \textit{complementiser}],\ 0.28 \right\rangle$ ($\boxed{2}$ is not a CP)
  - $\left\langle \begin{bmatrix} \text{PHON} & \langle \textit{depend} \rangle \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix} \rightarrow \neg \boxed{2}[\ldots|\text{HEAD } \textit{noun}],\ 0.93 \right\rangle$ ($\boxed{2}$ is not an NP)
- such constraints could be stated at the level of **word classes** rather than particular words (cf. Davis *et al.* 2024: 351), perhaps within more general linking rules.

Gradient selectional restrictions in HPSG 2

Solution 2 (more general, preferred):

- **lexical entries do not mention syntactic selectional restrictions** at all, only the number of arguments – or not even this, given some linking theory (Davis *et al.* 2024), e.g.:
- $\mathrm{LE}_{\mathit{depend}} \stackrel{\mathit{df}}{\equiv} \begin{bmatrix} \text{PHON} & \langle \mathit{depend} \rangle \\ \ldots & \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix}$
- then add **separate constraints** on particular realizations of particular arguments, e.g.:
  - $\left\langle \begin{bmatrix} \text{PHON} & \langle \mathit{depend} \rangle \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix} \rightarrow \neg \boxed{2}\left[\ldots \middle| \text{HEAD} \;\; \mathit{preposition}\right],\;\; 0.00 \right\rangle$ ($\boxed{2}$ is not a PP)
  - $\left\langle \begin{bmatrix} \text{PHON} & \langle \mathit{depend} \rangle \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix} \rightarrow \neg \boxed{2}\left[\ldots \middle| \text{HEAD} \;\; \mathit{complementiser}\right],\;\; 0.28 \right\rangle$ ($\boxed{2}$ is not a CP)
  - $\left\langle \begin{bmatrix} \text{PHON} & \langle \mathit{depend} \rangle \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix} \rightarrow \neg \boxed{2}\left[\ldots \middle| \text{HEAD} \;\; \mathit{noun}\right],\;\; 0.93 \right\rangle$ ($\boxed{2}$ is not an NP)
- such constraints could be stated at the level of **word classes** rather than particular words (cf. Davis *et al.* 2024: 351), perhaps within more general linking rules.

Gradient selectional restrictions in HPSG 2

Solution 2 (more general, preferred):

- **lexical entries do not mention syntactic selectional restrictions** at all, only the number of arguments – or not even this, given some linking theory (Davis *et al.* 2024), e.g.:
- $\mathrm{LE}_{\textit{depend}} \stackrel{df}{\equiv} \begin{bmatrix} \text{PHON} & \langle \textit{depend} \rangle \\ \ldots & \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix}$
- then add **separate constraints** on particular realizations of particular arguments, e.g.:
  - $\left\langle \begin{bmatrix} \text{PHON} & \langle \textit{depend} \rangle \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix} \rightarrow \neg \boxed{2}\left[\ldots \middle| \text{HEAD } \textit{preposition}\right],\ 0.00 \right\rangle$ ($\boxed{2}$ is not a PP)
  - $\left\langle \begin{bmatrix} \text{PHON} & \langle \textit{depend} \rangle \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix} \rightarrow \neg \boxed{2}\left[\ldots \middle| \text{HEAD } \textit{complementiser}\right],\ 0.28 \right\rangle$ ($\boxed{2}$ is not a CP)
  - $\left\langle \begin{bmatrix} \text{PHON} & \langle \textit{depend} \rangle \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix} \rightarrow \neg \boxed{2}\left[\ldots \middle| \text{HEAD } \textit{noun}\right],\ 0.93 \right\rangle$ ($\boxed{2}$ is not an NP)
- such constraints could be stated at the level of **word classes** rather than particular words (cf. Davis *et al.* 2024: 351), perhaps within more general linking rules.

Gradient selectional restrictions in HPSG 3

The **above constraints** are coarse-grained, at the level of general **syntactic categories**, e.g.:

- - $\left\langle \begin{bmatrix} \text{PHON} & \langle \textit{depend} \rangle \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix} \rightarrow \neg\boxed{2}\left[\ldots|\text{HEAD}\ \underline{\textit{preposition}}\right],\ 0.00 \right\rangle$ ($\boxed{2}$ is not a PP)
  - $\left\langle \begin{bmatrix} \text{PHON} & \langle \textit{depend} \rangle \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix} \rightarrow \neg\boxed{2}\left[\ldots|\text{HEAD}\ \underline{\textit{complementiser}}\right],\ 0.28 \right\rangle$ ($\boxed{2}$ is not a CP)

But not any PP or CP will do, e.g.:

- - **Justin Bieber can* ***depend*** [$_{PP}$ ***for*** *the fact* [$_{CP}$ *that his fans still love his early songs*]].
  - **Justin Bieber can* ***depend*** [$_{CP}$ ***if*** *his fans still love his early songs*].

Additional – fine-grained – **lexical and morphosyntactic constraints** are needed, e.g.:

- - $\left\langle \begin{bmatrix} \text{PHON} & \langle \textit{depend} \rangle \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2}\left[\ldots|\text{HEAD}\ \underline{\textit{preposition}}\right] \rangle \end{bmatrix} \rightarrow \boxed{2}\left[\ldots|\text{HEAD}|\underline{\text{PFORM}}\ \underline{\textit{on}}\right],\ 0.97 \right\rangle$

    (If $\boxed{2}$ is a PP, then it is a PP[*on*].)
  - $\left\langle \begin{bmatrix} \text{PHON} & \langle \textit{depend} \rangle \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2}\left[\ldots|\text{HEAD}\ \underline{\textit{complementiser}}\right] \rangle \end{bmatrix} \rightarrow \boxed{2}\left[\ldots|\text{HEAD}|\underline{\text{COMPFORM}}\ \underline{\textit{that}}\right],\ 0.91 \right\rangle$

    (If $\boxed{2}$ is a CP, then it is a CP[*that*].)

Again, such constraints could be stated at the level of **word classes**, within **linking rules**.

Gradient selectional restrictions in HPSG 3

The **above constraints** are coarse-grained, at the level of general **syntactic categories**, e.g.:

- - $\left\langle \begin{bmatrix} \text{PHON} & \langle \textit{depend} \rangle \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix} \rightarrow \neg\boxed{2}[\ldots|\text{HEAD}\ \underline{\textit{preposition}}],\ 0.00 \right\rangle$ ($\boxed{2}$ is not a PP)
  - $\left\langle \begin{bmatrix} \text{PHON} & \langle \textit{depend} \rangle \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix} \rightarrow \neg\boxed{2}[\ldots|\text{HEAD}\ \underline{\textit{complementiser}}],\ 0.28 \right\rangle$ ($\boxed{2}$ is not a CP)

But not any PP or CP will do, e.g.:

- - **Justin Bieber can* ***depend*** [$_{PP}$ ***for*** *the fact* [$_{CP}$ *that his fans still love his early songs*]].
  - **Justin Bieber can* ***depend*** [$_{CP}$ ***if*** *his fans still love his early songs*].

Additional – fine-grained – **lexical and morphosyntactic constraints** are needed, e.g.:

- - $\left\langle \begin{bmatrix} \text{PHON} & \langle \textit{depend} \rangle \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2}[\ldots|\text{HEAD}\ \underline{\textit{preposition}}] \rangle \end{bmatrix} \rightarrow \boxed{2}[\ldots|\text{HEAD}|\underline{\text{PFORM}}\ \underline{\textit{on}}],\ 0.97 \right\rangle$

    (If $\boxed{2}$ is a PP, then it is a PP[*on*].)
  - $\left\langle \begin{bmatrix} \text{PHON} & \langle \textit{depend} \rangle \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2}[\ldots|\text{HEAD}\ \underline{\textit{complementiser}}] \rangle \end{bmatrix} \rightarrow \boxed{2}[\ldots|\text{HEAD}|\underline{\text{COMPFORM}}\ \underline{\textit{that}}],\ 0.91 \right\rangle$

    (If $\boxed{2}$ is a CP, then it is a CP[*that*].)

Again, such constraints could be stated at the level of **word classes**, within **linking rules**.

Gradient selectional restrictions in HPSG 3

The **above constraints** are coarse-grained, at the level of general **syntactic categories**, e.g.:

- $\left\langle \begin{bmatrix} \text{PHON} & \langle \textit{depend} \rangle \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix} \rightarrow \neg \boxed{2} [\ldots|\text{HEAD}\ \underline{\textit{preposition}}],\ 0.00 \right\rangle$ ($\boxed{2}$ is not a PP)
- $\left\langle \begin{bmatrix} \text{PHON} & \langle \textit{depend} \rangle \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix} \rightarrow \neg \boxed{2} [\ldots|\text{HEAD}\ \underline{\textit{complementiser}}],\ 0.28 \right\rangle$ ($\boxed{2}$ is not a CP)

But not any PP or CP will do, e.g.:

- **Justin Bieber can **depend** [$_{\text{PP}}$ **for** the fact [$_{\text{CP}}$ that his fans still love his early songs]].*
- **Justin Bieber can **depend** [$_{\text{CP}}$ **if** his fans still love his early songs].*

Additional – fine-grained – **lexical and morphosyntactic constraints** are needed, e.g.:

- $\left\langle \begin{bmatrix} \text{PHON} & \langle \textit{depend} \rangle \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2} [\ldots|\text{HEAD}\ \underline{\textit{preposition}}] \rangle \end{bmatrix} \rightarrow \boxed{2} [\ldots|\text{HEAD}|\underline{\text{PFORM}}\ \underline{\textit{on}}],\ 0.97 \right\rangle$

  (If $\boxed{2}$ is a PP, then it is a PP[*on*].)
- $\left\langle \begin{bmatrix} \text{PHON} & \langle \textit{depend} \rangle \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2} [\ldots|\text{HEAD}\ \underline{\textit{complementiser}}] \rangle \end{bmatrix} \rightarrow \boxed{2} [\ldots|\text{HEAD}|\underline{\text{COMPFORM}}\ \underline{\textit{that}}],\ 0.91 \right\rangle$

  (If $\boxed{2}$ is a CP, then it is a CP[*that*].)

Again, such constraints could be stated at the level of **word classes**, within **linking rules**.

Gradient selectional restrictions in HPSG 3

The **above constraints** are coarse-grained, at the level of general **syntactic categories**, e.g.:

- $\left\langle \begin{bmatrix} \text{PHON} & \langle \textit{depend} \rangle \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix} \rightarrow \neg\boxed{2}[\ldots|\text{HEAD } \underline{\textit{preposition}}],\ 0.00 \right\rangle$ ($\boxed{2}$ is not a PP)
- $\left\langle \begin{bmatrix} \text{PHON} & \langle \textit{depend} \rangle \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix} \rightarrow \neg\boxed{2}[\ldots|\text{HEAD } \underline{\textit{complementiser}}],\ 0.28 \right\rangle$ ($\boxed{2}$ is not a CP)

But not any PP or CP will do, e.g.:

- **Justin Bieber can **depend** [$_{\text{PP}}$ **for** the fact [$_{\text{CP}}$ that his fans still love his early songs]].*
- **Justin Bieber can **depend** [$_{\text{CP}}$ **if** his fans still love his early songs].*

Additional – fine-grained – **lexical and morphosyntactic constraints** are needed, e.g.:

- $\left\langle \begin{bmatrix} \text{PHON} & \langle \textit{depend} \rangle \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2}[\ldots|\text{HEAD } \underline{\textit{preposition}}] \rangle \end{bmatrix} \rightarrow \boxed{2}[\ldots|\text{HEAD}|\underline{\text{PFORM}}\ \underline{\textit{on}}],\ 0.97 \right\rangle$

(If $\boxed{2}$ is a PP, then it is a PP[*on*].)

- $\left\langle \begin{bmatrix} \text{PHON} & \langle \textit{depend} \rangle \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2}[\ldots|\text{HEAD } \underline{\textit{complementiser}}] \rangle \end{bmatrix} \rightarrow \boxed{2}[\ldots|\text{HEAD}|\underline{\text{COMPFORM}}\ \underline{\textit{that}}],\ 0.91 \right\rangle$

(If $\boxed{2}$ is a CP, then it is a CP[*that*].)

Again, such constraints could be stated at the level of **word classes**, within **linking rules**.

Gradient selectional restrictions in HPSG 3

The **above constraints** are coarse-grained, at the level of general **syntactic categories**, e.g.:

- $\langle \begin{bmatrix} \text{PHON} & \langle \textit{depend} \rangle \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix} \rightarrow \neg\boxed{2}[\ldots|\text{HEAD}\ \underline{\textit{preposition}}],\ 0.00 \rangle$ ($\boxed{2}$ is not a PP)
- $\langle \begin{bmatrix} \text{PHON} & \langle \textit{depend} \rangle \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix} \rightarrow \neg\boxed{2}[\ldots|\text{HEAD}\ \underline{\textit{complementiser}}],\ 0.28 \rangle$ ($\boxed{2}$ is not a CP)

But not any PP or CP will do, e.g.:

- **Justin Bieber can **depend** [$_{PP}$ **for** the fact [$_{CP}$ that his fans still love his early songs]].*
- **Justin Bieber can **depend** [$_{CP}$ **if** his fans still love his early songs].*

Additional – fine-grained – **lexical and morphosyntactic constraints** are needed, e.g.:

- $\langle \begin{bmatrix} \text{PHON} & \langle \textit{depend} \rangle \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2}[\ldots|\text{HEAD}\ \underline{\textit{preposition}}] \rangle \end{bmatrix} \rightarrow \boxed{2}[\ldots|\text{HEAD}|\underline{\text{PFORM}}\ \underline{\textit{on}}],\ 0.97 \rangle$

  (If $\boxed{2}$ is a PP, then it is a PP[*on*].)
- $\langle \begin{bmatrix} \text{PHON} & \langle \textit{depend} \rangle \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2}[\ldots|\text{HEAD}\ \underline{\textit{complementiser}}] \rangle \end{bmatrix} \rightarrow \boxed{2}[\ldots|\text{HEAD}|\underline{\text{COMPFORM}}\ \underline{\textit{that}}],\ 0.91 \rangle$

  (If $\boxed{2}$ is a CP, then it is a CP[*that*].)

Again, such constraints could be stated at the level of **word classes**, within **linking rules**.

Gradient selectional restrictions in HPSG 3

The **above constraints** are coarse-grained, at the level of general **syntactic categories**, e.g.:

- - $\left\langle \begin{bmatrix} \text{PHON} & \langle \textit{depend} \rangle \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix} \rightarrow \neg\boxed{2}[\ldots|\text{HEAD}\ \underline{\textit{preposition}}],\ 0.00 \right\rangle$ (2 is not a PP)
  - $\left\langle \begin{bmatrix} \text{PHON} & \langle \textit{depend} \rangle \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix} \rightarrow \neg\boxed{2}[\ldots|\text{HEAD}\ \underline{\textit{complementiser}}],\ 0.28 \right\rangle$ (2 is not a CP)

But not any PP or CP will do, e.g.:

- - **Justin Bieber can **depend** [$_{PP}$ **for** the fact [$_{CP}$ that his fans still love his early songs]].*
  - **Justin Bieber can **depend** [$_{CP}$ **if** his fans still love his early songs].*

Additional – fine-grained – **lexical and morphosyntactic constraints** are needed, e.g.:

- - $\left\langle \begin{bmatrix} \text{PHON} & \langle \textit{depend} \rangle \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2}[\ldots|\text{HEAD}\ \underline{\textit{preposition}}] \rangle \end{bmatrix} \rightarrow \boxed{2}[\ldots|\text{HEAD}|\underline{\text{PFORM}}\ \underline{\textit{on}}],\ 0.97 \right\rangle$

    (If 2 is a PP, then it is a PP[*on*].)
  - $\left\langle \begin{bmatrix} \text{PHON} & \langle \textit{depend} \rangle \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2}[\ldots|\text{HEAD}\ \underline{\textit{complementiser}}] \rangle \end{bmatrix} \rightarrow \boxed{2}[\ldots|\text{HEAD}|\underline{\text{COMPFORM}}\ \underline{\textit{that}}],\ 0.91 \right\rangle$

    (If 2 is a CP, then it is a CP[*that*].)

Again, such constraints could be stated at the level of **word classes**, within **linking rules**.

Gradient selectional restrictions in HPSG 4

Summary of the **preferred solution**:

- **general category constraints** like:
  - $\left\langle \begin{bmatrix} \text{PHON} & \langle \textit{depend} \rangle \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix} \rightarrow \neg\boxed{2}\left[\ldots|\text{HEAD}\ \textit{complementiser}\right],\ 0.28 \right\rangle$
  - “The second argument of *depend* is not a CP (or pay 0.28).”
- **specific lexical and morphosyntactic constraints** like:
  - $\left\langle \begin{bmatrix} \text{PHON} & \langle \textit{depend} \rangle \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2}\left[\ldots|\text{HEAD}\ \textit{complementiser}\right] \rangle \end{bmatrix} \rightarrow \boxed{2}\left[\ldots|\text{HEAD}|\text{COMPFORM}\ \textit{that}\right],\ 0.91 \right\rangle$
  - “If the second argument of *depend* is a CP, then it is a CP[*that*] (or pay 0.91).”

Examples:

- ?*Justin Bieber can **depend*** [$_{CP}$ ***that*** *his fans still love his early songs*].
- 1st constraint violated, 2nd not – burden 0.28
- **Justin Bieber can **depend*** [$_{CP}$ ***if*** *his fans still love his early songs*].
- 1st constraint violated, 2nd also – burden $0.28 + 0.91 = 1.19$

Gradient selectional restrictions in HPSG 4

Summary of the **preferred solution**:

- **general category constraints** like:
  - $\left\langle \begin{bmatrix} \text{PHON} & \langle \textit{depend} \rangle \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix} \rightarrow \neg\boxed{2}[\ldots|\text{HEAD}\ \textit{complementiser}],\ 0.28 \right\rangle$
  - "The second argument of *depend* is not a CP (or pay 0.28)."
- **specific lexical and morphosyntactic constraints** like:
  - $\left\langle \begin{bmatrix} \text{PHON} & \langle \textit{depend} \rangle \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2}[\ldots|\text{HEAD}\ \textit{complementiser}] \rangle \end{bmatrix} \rightarrow \boxed{2}[\ldots|\text{HEAD}|\text{COMPFORM}\ \textit{that}],\ 0.91 \right\rangle$
  - "If the second argument of *depend* is a CP, then it is a CP[*that*] (or pay 0.91)."

Examples:

- ?*Justin Bieber can **depend*** [$_{CP}$ ***that*** *his fans still love his early songs*].
- 1st constraint violated, 2nd not – burden 0.28
- **Justin Bieber can **depend*** [$_{CP}$ ***if*** *his fans still love his early songs*].
- 1st constraint violated, 2nd also – burden $0.28 + 0.91 = 1.19$

Gradient selectional restrictions in HPSG 4

Summary of the **preferred solution**:

- **general category constraints** like:
  - $\left\langle \begin{bmatrix} \text{PHON} & \langle \textit{depend} \rangle \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix} \rightarrow \neg \boxed{2}\left[\ldots|\text{HEAD}\ \ \textit{complementiser}\right],\ 0.28 \right\rangle$
  - "The second argument of *depend* is not a CP (or pay 0.28)."
- **specific lexical and morphosyntactic constraints** like:
  - $\left\langle \begin{bmatrix} \text{PHON} & \langle \textit{depend} \rangle \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2}\left[\ldots|\text{HEAD}\ \ \textit{complementiser}\right] \rangle \end{bmatrix} \rightarrow \boxed{2}\left[\ldots|\text{HEAD}|\text{COMPFORM}\ \ \textit{that}\right],\ 0.91 \right\rangle$
  - "If the second argument of *depend* is a CP, then it is a CP[*that*] (or pay 0.91)."

Examples:

- ?*Justin Bieber can* ***depend*** [$_{\text{CP}}$ ***that*** *his fans still love his early songs*].
- 1st constraint violated, 2nd not – burden 0.28
- **Justin Bieber can* ***depend*** [$_{\text{CP}}$ ***if*** *his fans still love his early songs*].
- 1st constraint violated, 2nd also – burden $0.28 + 0.91 = 1.19$

Gradient selectional restrictions in HPSG 4

Summary of the **preferred solution**:

- **general category constraints** like:
  - $\left\langle \begin{bmatrix} \text{PHON} & \langle \textit{depend} \rangle \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix} \rightarrow \neg\boxed{2}\big[\ldots|\text{HEAD}\ \ \textit{complementiser}\big],\ 0.28 \right\rangle$
  - "The second argument of *depend* is not a CP (or pay 0.28)."
- **specific lexical and morphosyntactic constraints** like:
  - $\left\langle \begin{bmatrix} \text{PHON} & \langle \textit{depend} \rangle \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2}\big[\ldots|\text{HEAD}\ \ \textit{complementiser}\big] \rangle \end{bmatrix} \rightarrow \boxed{2}\big[\ldots|\text{HEAD}|\text{COMPFORM}\ \ \textit{that}\big],\ 0.91 \right\rangle$
  - "If the second argument of *depend* is a CP, then it is a CP[*that*] (or pay 0.91)."

Examples:

- ?*Justin Bieber can **depend** [$_{CP}$ **that** his fans still love his early songs].*
- 1st constraint violated, 2nd not – burden 0.28
- **Justin Bieber can **depend** [$_{CP}$ **if** his fans still love his early songs].*
- 1st constraint violated, 2nd also – burden $0.28 + 0.91 = 1.19$

Gradient selectional restrictions in HPSG 4

Summary of the **preferred solution**:

- **general category constraints** like:
  - $\left\langle \begin{bmatrix} \text{PHON} & \langle \textit{depend} \rangle \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix} \rightarrow \neg\boxed{2}\left[\ldots|\text{HEAD}\ \ \textit{complementiser}\right],\ 0.28 \right\rangle$
  - "The second argument of *depend* is not a CP (or pay 0.28)."
- **specific lexical and morphosyntactic constraints** like:
  - $\left\langle \begin{bmatrix} \text{PHON} & \langle \textit{depend} \rangle \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2}\left[\ldots|\text{HEAD}\ \ \textit{complementiser}\right] \rangle \end{bmatrix} \rightarrow \boxed{2}\left[\ldots|\text{HEAD}|\text{COMPFORM}\ \ \textit{that}\right],\ 0.91 \right\rangle$
  - "If the second argument of *depend* is a CP, then it is a CP[*that*] (or pay 0.91)."

Examples:

- ?*Justin Bieber can* ***depend*** [$_{CP}$ ***that*** *his fans still love his early songs*].
- 1st constraint violated, 2nd not – burden 0.28
- **Justin Bieber can* ***depend*** [$_{CP}$ ***if*** *his fans still love his early songs*].
- 1st constraint violated, 2nd also – burden $0.28 + 0.91 = 1.19$

Gradient selectional restrictions in HPSG 4

Summary of the **preferred solution**:

- **general category constraints** like:
  - $\left\langle \begin{bmatrix} \text{PHON} & \langle \textit{depend} \rangle \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix} \rightarrow \neg\boxed{2}\left[\ldots|\text{HEAD}\ \textit{complementiser}\right],\ 0.28 \right\rangle$
  - "The second argument of *depend* is not a CP (or pay 0.28)."
- **specific lexical and morphosyntactic constraints** like:
  - $\left\langle \begin{bmatrix} \text{PHON} & \langle \textit{depend} \rangle \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2}\left[\ldots|\text{HEAD}\ \textit{complementiser}\right] \rangle \end{bmatrix} \rightarrow \boxed{2}\left[\ldots|\text{HEAD}|\text{COMPFORM}\ \textit{that}\right],\ 0.91 \right\rangle$
  - "If the second argument of *depend* is a CP, then it is a CP[*that*] (or pay 0.91)."

Examples:

- ? *Justin Bieber can* ***depend*** [$_{\text{CP}}$ ***that*** *his fans still love his early songs*].
- 1st constraint violated, 2nd not – burden 0.28
- * *Justin Bieber can* ***depend*** [$_{\text{CP}}$ ***if*** *his fans still love his early songs*].
- 1st constraint violated, 2nd also – burden $0.28 + 0.91 = 1.19$

Gradient selectional restrictions in HPSG 4

Summary of the **preferred solution**:

- **general category constraints** like:
  - $\left\langle \begin{bmatrix} \text{PHON} & \langle \textit{depend} \rangle \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix} \rightarrow \neg\boxed{2}[\ldots|\text{HEAD}\ \textit{complementiser}],\ 0.28 \right\rangle$
  - "The second argument of *depend* is not a CP (or pay 0.28)."
- **specific lexical and morphosyntactic constraints** like:
  - $\left\langle \begin{bmatrix} \text{PHON} & \langle \textit{depend} \rangle \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2}[\ldots|\text{HEAD}\ \textit{complementiser}] \rangle \end{bmatrix} \rightarrow \boxed{2}[\ldots|\text{HEAD}|\text{COMPFORM}\ \textit{that}],\ 0.91 \right\rangle$
  - "If the second argument of *depend* is a CP, then it is a CP[*that*] (or pay 0.91)."

Examples:

- ? *Justin Bieber can **depend** [$_{CP}$ **that** his fans still love his early songs].*
- 1st constraint violated, 2nd not – burden 0.28
- * *Justin Bieber can **depend** [$_{CP}$ **if** his fans still love his early songs].*
- 1st constraint violated, 2nd also – burden 0.28 + 0.91 = 1.19

Gradient selectional restrictions in HPSG 4

Summary of the **preferred solution**:

- **general category constraints** like:
  - $\left\langle \begin{bmatrix} \text{PHON} & \langle \textit{depend} \rangle \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix} \rightarrow \neg\boxed{2}[\ldots|\text{HEAD}\ \textit{complementiser}],\ 0.28 \right\rangle$
  - "The second argument of *depend* is not a CP (or pay 0.28)."
- **specific lexical and morphosyntactic constraints** like:
  - $\left\langle \begin{bmatrix} \text{PHON} & \langle \textit{depend} \rangle \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2}[\ldots|\text{HEAD}\ \textit{complementiser}] \rangle \end{bmatrix} \rightarrow \boxed{2}[\ldots|\text{HEAD}|\text{COMPFORM}\ \textit{that}],\ 0.91 \right\rangle$
  - "If the second argument of *depend* is a CP, then it is a CP[*that*] (or pay 0.91)."

Examples:

- ? *Justin Bieber can **depend** [$_{CP}$ **that** his fans still love his early songs].*
- 1st constraint violated, 2nd not – burden 0.28
- * *Justin Bieber can **depend** [$_{CP}$ **if** his fans still love his early songs].*
- 1st constraint violated, 2nd also – burden $0.28 + 0.91 = 1.19$

Meaning of weights

What is the **meaning of weights**?

Ideally:

- they are **neurocognitively real**, i.e.,
- they **directly reflect** the magnitude of neurocognitive **responses to constraint violations within an individual**,
- so there may be individual differences;
- **not clear how to reliably estimate such individual responses**.

Practically:

- the **weights reflect results of a particular experiment** on a population sample,
- **averaging individual responses** in the sense above.

Meaning of weights

What is the **meaning of weights**?

Ideally:

- they are **neurocognitively real**, i.e.,
- they **directly reflect** the magnitude of neurocognitive **responses to constraint violations within an individual**,
- so there may be individual differences;
- **not clear how to reliably estimate such individual responses**.

Practically:

- the **weights reflect results of a particular experiment** on a population sample,
- **averaging individual responses** in the sense above.

Meaning of weights

What is the **meaning of weights**?

Ideally:

- they are **neurocognitively real**, i.e.,
- they **directly reflect** the magnitude of neurocognitive **responses to constraint violations** **within an individual**,
- so there may be individual differences;
- **not clear how to reliably estimate such individual responses.**

Practically:

- the **weights** **reflect results of a particular experiment** on a population sample,
- **averaging individual responses** in the sense above.

Meaning of weights

What is the **meaning of weights**?

Ideally:

- they are **neurocognitively real**, i.e.,
- they **directly reflect** the magnitude of neurocognitive **responses to constraint violations within an individual**,
- so there may be individual differences;
- not clear how to reliably estimate such individual responses.

Practically:

- the **weights reflect results of a particular experiment** on a population sample,
- **averaging individual responses** in the sense above.

Meaning of weights

What is the **meaning of weights**?

Ideally:

- they are **neurocognitively real**, i.e.,
- they **directly reflect** the magnitude of neurocognitive **responses to constraint violations within an individual**,
- so there may be individual differences;
- **not clear how to reliably estimate such individual responses**.

Practically:

- the **weights reflect results of a particular experiment** on a population sample,
- **averaging individual responses** in the sense above.

Meaning of weights

What is the **meaning of weights**?

Ideally:

- they are **neurocognitively real**, i.e.,
- they **directly reflect** the magnitude of neurocognitive **responses to constraint violations within an individual**,
- so there may be individual differences;
- **not clear how to reliably estimate such individual responses**.

Practically:

- the **weights reflect results of a particular experiment** on a population sample,
- **averaging individual responses** in the sense above.

Meaning of weights

What is the **meaning of weights**?

Ideally:

- they are **neurocognitively real**, i.e.,
- they **directly reflect** the magnitude of neurocognitive **responses to constraint violations within an individual**,
- so there may be individual differences;
- **not clear how to reliably estimate such individual responses**.

Practically:

- the **weights reflect results of a particular experiment** on a population sample,
- **averaging individual responses** in the sense above.

Normalizing weights

Ideally, different experimental methodologies (Likert Scale, Magnitude Estimation, Thermometer Method, etc.) should result in similar weights.

So there should be a **normalizing procedure** mapping specific experimental results to a normalized scale of weights, e.g., $[0, 1]$.

Here, **the simplest normalizing procedure** (to be verified by future experiments):

- **suppose** that:
  - the **experimental scale** is $[S_{min}, S_{max}]$ (e.g., $[1, 7]$ Likert Scale),
  - the **estimated judgement drop** caused by a violation of constraint δ_i is β_i,
  - then the **weight** of this constraint is $w_i = \frac{|\beta_i|}{S_{max} - S_{min}}$;
- **in the experiment above**:
  - the **experimental scale** is (practically) $[10, 40]$ (Thermometer Method),
  - the mixed effects model coefficient reflecting **the drop from the grammatical PP condition to the CP condition** for *depend* is 8.44,
  - so the **weight** of the constraint prohibiting CPs as second arguments of *depend* is: $w_{\text{DEPEND\_CP}} = \frac{8.44}{40-10} = 0.28$

Normalizing weights

Ideally, different experimental methodologies (Likert Scale, Magnitude Estimation, Thermometer Method, etc.) should result in similar weights.

So there should be a **normalizing procedure** mapping specific experimental results to a normalized scale of weights, e.g., $[0, 1]$.

Here, **the simplest normalizing procedure** (to be verified by future experiments):

- **suppose** that:
  - the **experimental scale** is $[S_{min}, S_{max}]$ (e.g., $[1, 7]$ Likert Scale),
  - the **estimated judgement drop** caused by a violation of constraint δ_i is β_i,
  - then the **weight** of this constraint is $w_i = \frac{|\beta_i|}{S_{max} - S_{min}}$;
- **in the experiment above**:
  - the **experimental scale** is (practically) $[10, 40]$ (Thermometer Method),
  - the mixed effects model coefficient reflecting **the drop from the grammatical PP condition to the CP condition** for *depend* is 8.44,
  - so the **weight** of the constraint prohibiting CPs as second arguments of *depend* is:

$$w_{\text{DEPEND\_CP}} = \frac{8.44}{40-10} = 0.28$$

Normalizing weights

Ideally, different experimental methodologies (Likert Scale, Magnitude Estimation, Thermometer Method, etc.) should result in similar weights.

So there should be a **normalizing procedure** mapping specific experimental results to a normalized scale of weights, e.g., $[0, 1]$.

Here, **the simplest normalizing procedure** (to be verified by future experiments):

- **suppose** that:
  - the **experimental scale** is $[S_{min}, S_{max}]$ (e.g., $[1, 7]$ Likert Scale),
  - the **estimated judgement drop** caused by a violation of constraint δ_i is β_i,
  - then the **weight** of this constraint is $w_i = \frac{|\beta_i|}{S_{max} - S_{min}}$;
- **in the experiment above**:
  - the **experimental scale** is (practically) $[10, 40]$ (Thermometer Method),
  - the mixed effects model coefficient reflecting **the drop from the grammatical PP condition to the CP condition** for *depend* is 8.44,
  - so the **weight** of the constraint prohibiting CPs as second arguments of *depend* is: $w_{\text{DEPEND\_CP}} = \frac{8.44}{40-10} = 0.28$

Normalizing weights

Ideally, different experimental methodologies (Likert Scale, Magnitude Estimation, Thermometer Method, etc.) should result in similar weights.

So there should be a **normalizing procedure** mapping specific experimental results to a normalized scale of weights, e.g., $[0, 1]$.

Here, **the simplest normalizing procedure** (to be verified by future experiments):

- **suppose** that:
  - the **experimental scale** is $[S_{min}, S_{max}]$ (e.g., $[1, 7]$ Likert Scale),
  - the **estimated judgement drop** caused by a violation of constraint δ_i is β_i,
  - then the **weight** of this constraint is $w_i = \frac{|\beta_i|}{S_{max} - S_{min}}$;
- **in the experiment above**:
  - the **experimental scale** is (practically) $[10, 40]$ (Thermometer Method),
  - the mixed effects model coefficient reflecting **the drop from the grammatical PP condition to the CP condition** for *depend* is 8.44,
  - so the **weight** of the constraint prohibiting CPs as second arguments of *depend* is:

    $w_{\text{DEPEND\_CP}} = \frac{8.44}{40-10} = 0.28$

Normalizing weights

Ideally, different experimental methodologies (Likert Scale, Magnitude Estimation, Thermometer Method, etc.) should result in similar weights.

So there should be a **normalizing procedure** mapping specific experimental results to a normalized scale of weights, e.g., $[0, 1]$.

Here, **the simplest normalizing procedure** (to be verified by future experiments):

- **suppose** that:
  - the **experimental scale** is $[S_{min}, S_{max}]$ (e.g., $[1, 7]$ Likert Scale),
  - the **estimated judgement drop** caused by a violation of constraint δ_i is β_i,
  - then the **weight** of this constraint is $w_i = \frac{|\beta_i|}{S_{max} - S_{min}}$;
- **in the experiment above**:
  - the **experimental scale** is (practically) $[10, 40]$ (Thermometer Method),
  - the mixed effects model coefficient reflecting **the drop from the grammatical PP condition to the CP condition** for *depend* is 8.44,
  - so the **weight** of the constraint prohibiting CPs as second arguments of *depend* is: $w_{\text{DEPEND\_CP}} = \frac{8.44}{40-10} = 0.28$

Normalizing weights

Ideally, different experimental methodologies (Likert Scale, Magnitude Estimation, Thermometer Method, etc.) should result in similar weights.

So there should be a **normalizing procedure** mapping specific experimental results to a normalized scale of weights, e.g., $[0,1]$.

Here, **the simplest normalizing procedure** (to be verified by future experiments):

- **suppose** that:
  - the **experimental scale** is $[S_{min}, S_{max}]$ (e.g., $[1,7]$ Likert Scale),
  - the **estimated judgement drop** caused by a violation of constraint δ_i is β_i,
  - then the **weight** of this constraint is $w_i = \frac{|\beta_i|}{S_{max} - S_{min}}$;
- in the experiment above:
  - the **experimental scale** is (practically) $[10,40]$ (Thermometer Method),
  - the mixed effects model coefficient reflecting **the drop from the grammatical PP condition to the CP condition** for *depend* is 8.44,
  - so the **weight** of the constraint prohibiting CPs as second arguments of *depend* is:

$$w_{\text{DEPEND\_CP}} = \frac{8.44}{40-10} = 0.28$$

Normalizing weights

Ideally, different experimental methodologies (Likert Scale, Magnitude Estimation, Thermometer Method, etc.) should result in similar weights.

So there should be a **normalizing procedure** mapping specific experimental results to a normalized scale of weights, e.g., $[0, 1]$.

Here, **the simplest normalizing procedure** (to be verified by future experiments):

- **suppose** that:
  - the **experimental scale** is $[S_{min}, S_{max}]$ (e.g., $[1, 7]$ Likert Scale),
  - the **estimated judgement drop** caused by a violation of constraint δ_i is β_i,
  - then the **weight** of this constraint is $w_i = \frac{|\beta_i|}{S_{max} - S_{min}}$;
- **in the experiment above**:
  - the **experimental scale** is (practically) $[10, 40]$ (Thermometer Method),
  - the mixed effects model coefficient reflecting **the drop from the grammatical PP condition to the CP condition** for *depend* is 8.44,
  - so the **weight** of the constraint prohibiting CPs as second arguments of *depend* is: $w_{\text{DEPEND\_CP}} = \frac{8.44}{40-10} = 0.28$

Normalizing weights

Ideally, different experimental methodologies (Likert Scale, Magnitude Estimation, Thermometer Method, etc.) should result in similar weights.

So there should be a **normalizing procedure** mapping specific experimental results to a normalized scale of weights, e.g., $[0, 1]$.

Here, **the simplest normalizing procedure** (to be verified by future experiments):

- **suppose** that:
  - the **experimental scale** is $[S_{min}, S_{max}]$ (e.g., $[1, 7]$ Likert Scale),
  - the **estimated judgement drop** caused by a violation of constraint δ_i is β_i,
  - then the **weight** of this constraint is $w_i = \frac{|\beta_i|}{S_{max} - S_{min}}$;
- **in the experiment above**:
  - the **experimental scale** is (practically) $[10, 40]$ (Thermometer Method),
  - the mixed effects model coefficient reflecting **the drop from the grammatical PP condition to the CP condition** for *depend* is 8.44,
  - so the **weight** of the constraint prohibiting CPs as second arguments of *depend* is: $w_{\text{DEPEND\_CP}} = \frac{8.44}{40-10} = 0.28$

Normalizing weights

Ideally, different experimental methodologies (Likert Scale, Magnitude Estimation, Thermometer Method, etc.) should result in similar weights.

So there should be a **normalizing procedure** mapping specific experimental results to a normalized scale of weights, e.g., $[0, 1]$.

Here, **the simplest normalizing procedure** (to be verified by future experiments):

- **suppose** that:
  - the **experimental scale** is $[S_{min}, S_{max}]$ (e.g., $[1, 7]$ Likert Scale),
  - the **estimated judgement drop** caused by a violation of constraint δ_i is β_i,
  - then the **weight** of this constraint is $w_i = \frac{|\beta_i|}{S_{max} - S_{min}}$;
- **in the experiment above**:
  - the **experimental scale** is (practically) $[10, 40]$ (Thermometer Method),
  - the mixed effects model coefficient reflecting **the drop from the grammatical PP condition to the CP condition** for *depend* is 8.44,
  - so the **weight** of the constraint prohibiting CPs as second arguments of *depend* is:

    $w_{\text{DEPEND\_CP}} = \frac{8.44}{40-10} = 0.28$

Summary and conclusion

A number of researchers have argued for **gradient grammaticality**, e.g., Chomsky 1965, Keller 2000, Sorace and Keller 2005, Featherston 2005, Häussler and Juzek 2016, Lau *et al.* 2017, . . . , Przepiórkowski *et al.* 2025; see also Francis 2022.

Our experiments **contribute to this mounting evidence**:

- predicates readily combining with PPs expressing propositions ("P the fact that...")
- differ in their acceptance of CPs (primary vessels of propositions).

Our **technical HPSG contribution**:

- a general approach to gradience in HPSG (based on Şenşekerci 2024),
- a specific approach to **gradient syntactic selectional restrictions**.

Also:

- a way to **extract normalized weights from experimental results**;
- contrast with approaches where weights reflect corpus frequencies (e.g., Probabilistic CFGs, Brew 1995, etc.),
- and see, e.g., Featherston 2005 on the differences between the two.

Summary and conclusion

A number of researchers have argued for **gradient grammaticality**, e.g., Chomsky 1965, Keller 2000, Sorace and Keller 2005, Featherston 2005, Häussler and Juzek 2016, Lau *et al.* 2017, …, Przepiórkowski *et al.* 2025; see also Francis 2022.

Our experiments **contribute to this mounting evidence**:

- predicates readily combining with PPs expressing propositions (“P the fact that…”)
- differ in their acceptance of CPs (primary vessels of propositions).

Our **technical HPSG contribution**:

- a general approach to gradience in HPSG (based on Şenşekerci 2024),
- a specific approach to **gradient syntactic selectional restrictions**.

Also:

- a way to **extract normalized weights from experimental results**;
- contrast with approaches where weights reflect corpus frequencies (e.g., Probabilistic CFGs, Brew 1995, etc.),
- and see, e.g., Featherston 2005 on the differences between the two.

Summary and conclusion

A number of researchers have argued for **gradient grammaticality**, e.g., Chomsky 1965, Keller 2000, Sorace and Keller 2005, Featherston 2005, Häussler and Juzek 2016, Lau *et al.* 2017, . . . , Przepiórkowski *et al.* 2025; see also Francis 2022.

Our experiments **contribute to this mounting evidence**:

- predicates readily combining with PPs expressing propositions ("P the fact that. . . ")
- differ in their acceptance of CPs (primary vessels of propositions).

Our **technical HPSG contribution**:

- a general approach to gradience in HPSG (based on Şenşekerci 2024),
- a specific approach to **gradient syntactic selectional restrictions**.

Also:

- a way to **extract normalized weights from experimental results**;
- contrast with approaches where weights reflect corpus frequencies (e.g., Probabilistic CFGs, Brew 1995, etc.),
- and see, e.g., Featherston 2005 on the differences between the two.

Summary and conclusion

A number of researchers have argued for **gradient grammaticality**, e.g., Chomsky 1965, Keller 2000, Sorace and Keller 2005, Featherston 2005, Häussler and Juzek 2016, Lau *et al.* 2017, . . . , Przepiórkowski *et al.* 2025; see also Francis 2022.

Our experiments **contribute to this mounting evidence**:

- predicates readily combining with PPs expressing propositions (“P the fact that. . . ”)
- differ in their acceptance of CPs (primary vessels of propositions).

Our **technical HPSG contribution**:

- a general approach to gradience in HPSG (based on Şenşekerci 2024),
- a specific approach to **gradient syntactic selectional restrictions**.

Also:

- a way to **extract normalized weights from experimental results**;
- contrast with approaches where weights reflect corpus frequencies (e.g., Probabilistic CFGs, Brew 1995, etc.),
- and see, e.g., Featherston 2005 on the differences between the two.

Summary and conclusion

A number of researchers have argued for **gradient grammaticality**, e.g., Chomsky 1965, Keller 2000, Sorace and Keller 2005, Featherston 2005, Häussler and Juzek 2016, Lau *et al.* 2017, . . . , Przepiórkowski *et al.* 2025; see also Francis 2022.

Our experiments **contribute to this mounting evidence**:

- predicates readily combining with PPs expressing propositions ("P the fact that. . . ")
- differ in their acceptance of CPs (primary vessels of propositions).

Our **technical HPSG contribution**:

- a general approach to gradience in HPSG (based on Şenşekerci 2024),
- a specific approach to **gradient syntactic selectional restrictions**.

Also:

- a way to **extract normalized weights from experimental results**;
- contrast with approaches where weights reflect corpus frequencies (e.g., Probabilistic CFGs, Brew 1995, etc.),
- and see, e.g., Featherston 2005 on the differences between the two.

Summary and conclusion

A number of researchers have argued for **gradient grammaticality**, e.g., Chomsky 1965, Keller 2000, Sorace and Keller 2005, Featherston 2005, Häussler and Juzek 2016, Lau *et al.* 2017, . . . , Przepiórkowski *et al.* 2025; see also Francis 2022.

Our experiments **contribute to this mounting evidence**:

- predicates readily combining with PPs expressing propositions ("P the fact that. . . ")
- differ in their acceptance of CPs (primary vessels of propositions).

Our **technical HPSG contribution**:

- a general approach to gradience in HPSG (based on Şenşekerci 2024),
- a specific approach to **gradient syntactic selectional restrictions**.

Also:

- a way to **extract normalized weights from experimental results**;
- contrast with approaches where weights reflect corpus frequencies (e.g., Probabilistic CFGs, Brew 1995, etc.),
- and see, e.g., Featherston 2005 on the differences between the two.

Summary and conclusion

A number of researchers have argued for **gradient grammaticality**, e.g., Chomsky 1965, Keller 2000, Sorace and Keller 2005, Featherston 2005, Häussler and Juzek 2016, Lau *et al.* 2017, . . . , Przepiórkowski *et al.* 2025; see also Francis 2022.

Our experiments **contribute to this mounting evidence**:

- predicates readily combining with PPs expressing propositions ("P the fact that. . . ")
- differ in their acceptance of CPs (primary vessels of propositions).

Our **technical HPSG contribution**:

- a general approach to gradience in HPSG (based on Şenşekerci 2024),
- a specific approach to **gradient syntactic selectional restrictions**.

Also:

- a way to **extract normalized weights from experimental results**;
- contrast with approaches where weights reflect corpus frequencies (e.g., Probabilistic CFGs, Brew 1995, etc.),
- and see, e.g., Featherston 2005 on the differences between the two.

Summary and conclusion

A number of researchers have argued for **gradient grammaticality**, e.g., Chomsky 1965, Keller 2000, Sorace and Keller 2005, Featherston 2005, Häussler and Juzek 2016, Lau *et al.* 2017, …, Przepiórkowski *et al.* 2025; see also Francis 2022.

Our experiments **contribute to this mounting evidence**:

- predicates readily combining with PPs expressing propositions ("P the fact that…")
- differ in their acceptance of CPs (primary vessels of propositions).

Our **technical HPSG contribution**:

- a general approach to gradience in HPSG (based on Şenşekerci 2024),
- a specific approach to **gradient syntactic selectional restrictions**.

Also:

- a way to **extract normalized weights from experimental results**;
- contrast with approaches where weights reflect corpus frequencies (e.g., Probabilistic CFGs, Brew 1995, etc.),
- and see, e.g., Featherston 2005 on the differences between the two.

Summary and conclusion

A number of researchers have argued for **gradient grammaticality**, e.g., Chomsky 1965, Keller 2000, Sorace and Keller 2005, Featherston 2005, Häussler and Juzek 2016, Lau *et al.* 2017, ..., Przepiórkowski *et al.* 2025; see also Francis 2022.

Our experiments **contribute to this mounting evidence**:

- predicates readily combining with PPs expressing propositions ("P the fact that...")
- differ in their acceptance of CPs (primary vessels of propositions).

Our **technical HPSG contribution**:

- a general approach to gradience in HPSG (based on Şenşekerci 2024),
- a specific approach to **gradient syntactic selectional restrictions**.

Also:

- a way to **extract normalized weights from experimental results**;
- contrast with approaches where weights reflect corpus frequencies (e.g., Probabilistic CFGs, Brew 1995, etc.),
- and see, e.g., Featherston 2005 on the differences between the two.

Summary and conclusion

A number of researchers have argued for **gradient grammaticality**, e.g., Chomsky 1965, Keller 2000, Sorace and Keller 2005, Featherston 2005, Häussler and Juzek 2016, Lau *et al.* 2017, ..., Przepiórkowski *et al.* 2025; see also Francis 2022.

Our experiments **contribute to this mounting evidence**:

- predicates readily combining with PPs expressing propositions ("P the fact that...")
- differ in their acceptance of CPs (primary vessels of propositions).

Our **technical HPSG contribution**:

- a general approach to gradience in HPSG (based on Şenşekerci 2024),
- a specific approach to **gradient syntactic selectional restrictions**.

Also:

- a way to **extract normalized weights from experimental results**;
- contrast with approaches where weights reflect corpus frequencies (e.g., Probabilistic CFGs, Brew 1995, etc.),
- and see, e.g., Featherston 2005 on the differences between the two.

Limitations and outlook

There are various **limitations** of this early-stage work, including:

- we assumed that certain selectional restrictions are syntactic, but **distinguishing between semantic and syntactic selectional restrictions is often difficult**;
- **selectional restrictions should probably be stated at the level of word classes** rather than particular words,
- and they should be **integrated with a linking theory**.

Still, we hope to have made some **progress in opening HPSG to gradient grammaticality**.

Limitations and outlook

There are various **limitations** of this early-stage work, including:

- we assumed that certain selectional restrictions are syntactic, but **distinguishing between semantic and syntactic selectional restrictions is often difficult**;
- **selectional restrictions should probably be stated at the level of word classes** rather than particular words,
- and they should be **integrated with a linking theory**.

Still, we hope to have made some **progress in opening HPSG to gradient grammaticality**.

Limitations and outlook

There are various **limitations** of this early-stage work, including:

- we assumed that certain selectional restrictions are syntactic, but **distinguishing between semantic and syntactic selectional restrictions is often difficult**;
- **selectional restrictions should probably be stated at the level of word classes** rather than particular words,
- and they should be **integrated with a linking theory**.

Still, we hope to have made some **progress in opening HPSG to gradient grammaticality**.

Limitations and outlook

There are various **limitations** of this early-stage work, including:

- we assumed that certain selectional restrictions are syntactic, but **distinguishing between semantic and syntactic selectional restrictions is often difficult**;
- **selectional restrictions should probably be stated at the level of word classes** rather than particular words,
- and they should be **integrated with a linking theory**.

Still, we hope to have made some **progress in opening HPSG to gradient grammaticality**.

Limitations and outlook

There are various **limitations** of this early-stage work, including:

- we assumed that certain selectional restrictions are syntactic, but **distinguishing between semantic and syntactic selectional restrictions is often difficult**;
- **selectional restrictions should probably be stated at the level of word classes** rather than particular words,
- and they should be **integrated with a linking theory**.

Still, we hope to have made some **progress in opening HPSG to gradient grammaticality**.

Limitations and outlook

There are various **limitations** of this early-stage work, including:

- we assumed that certain selectional restrictions are syntactic, but **distinguishing between semantic and syntactic selectional restrictions is often difficult**;
- **selectional restrictions should probably be stated at the level of word classes** rather than particular words,
- and they should be **integrated with a linking theory**.

Still, we hope to have made some **progress in opening HPSG to gradient grammaticality**.

Thank you for your attention!

Motto

*The current **disconnect between theoretical investigations** of language structure **and psycholinguistic studies** is an unfortunate feature of our discipline. Because **HPSG** comports so well with what is known about processing, it **could serve as the basis for a reconnection between these two areas of study**.*

(Wasow 2024: 1169)

Motto

*The current **disconnect between theoretical investigations** of language structure **and psycholinguistic studies** is an unfortunate feature of our discipline. Because HPSG comports so well with what is known about processing, it **could serve as the basis for a reconnection between these two areas of study**.*

(Wasow 2024: 1169)

Brew, C. (1995). Stochastic HPSG. In *Proceedings of the 7th Conference of the European Chapter of the Association for Computational Linguistics (EACL 1995)*, pp. 83–89, Dublin.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. MIT Press, Cambridge, MA.

Davis, A., Koenig, J.-P., and Wechsler, S. (2024). Argument structure and linking. In Müller *et al.* (2024), pp. 335–390.

Featherston, S. (2005). The Decathlon model of empirical syntax. In S. Kepser and M. Reis, eds., *Linguistic Evidence: Empirical, Theoretical and Computational Perspectives*, pp. 187–208. Mouton de Gruyter, Berlin/New York.

Featherston, S. (2008). Thermometer judgements as linguistic evidence. In C. M. Riehl and A. Rothe, eds., *Was ist linguistische Evidenz?*, pp. 69–90. Shaker Verlag, Aachen.

Featherston, S. (2009). A scale for measuring well-formedness: Why syntax needs boiling and freezing points. In S. Featherston and S. Winkler, eds., *The Fruits of Empirical Linguistics (Volume 1: Process)*, pp. 47–73. Mouton de Gruyter, Berlin.

Francis, E. J. (2022). *Gradient Acceptability and Linguistic Theory*. Oxford University Press, Oxford.

Gerbrich, H., Schreier, V., and Featherston, S. (2019). Standard items for English judgment studies: Syntax and semantics. In S. Featherston, R. Hörnig, S. von Wietersheim, and S. Winkler, eds., *Experiments in Focus: Information Structure and Semantic Processing*, pp. 305–327. Mouton de Gruyter, Berlin.

Häussler, J. and Juzek, T. S. (2016). Detecting and discouraging non-cooperative behavior in online experiments using an acceptability judgment task. In H. Christ, D. Klenovšak, L. Sönning, and V. Werner, eds., *A Blend of MaLT: Selected Contributions from the Methods and Linguistic*

Theories Symposium 2015, pp. 73–99. University of Bamberg Press.

Höhle, T. N. (1994). Spurenlose extraction. Unpubl. ms., University of Tübingen, reprinted in Müller *et al.* 2019.

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

Lau, J. H., Clark, A., and Lappin, S. (2017). Grammaticality, acceptability, and probability: A probabilistic view of linguistic knowledge. *Cognitive Science*, **41**, 1202–1241.

Müller, S., Reis, M., and Richter, F., eds. (2019). *Beiträge zur deutschen Grammatik: Gesammelte Schriften von Tilman N. Höhle*. Language Science Press, Berlin.

Müller, S., Abeillé, A., Borsley, R. D., and Koenig, J.-P., eds. (2024). *Head-Driven Phrase Structure Grammar: The Handbook*. Language Science Press, Berlin, 2nd edition.

Przepiórkowski, A., Łukasiewicz-Pater, J., Kuś, K., and Maćkiewicz, B. (2025). Heterofunctional coordination in German. *Journal of Comparative Germanic Linguistics*, **28**(3), 1–75.

Richter, F. (2004). *A Mathematical Formalism for Linguistic Theories with an Application in Head-Driven Phrase Structure Grammar*. Ph.D. dissertation (defended in 2000), Universität Tübingen.

Richter, F. (2024). Formal background. In Müller *et al.* (2024), pp. 93–131.

Şenşekerci, B. (2024). Gradient HPSG. In S. Müller and R. Chaves, eds., *Proceedings of the HPSG 2024 Conference*, pp. 173–192.

Sorace, A. and Keller, F. (2005). Gradience in linguistic data. *Lingua*, **115**(11), 1497–1524.

Wasow, T. (2024). Processing. In Müller *et al.* (2024), pp. 1153–1180.

Experimental Evidence 4

Another experiment (*familiar, depend, speak, talk*).

Acceptability judgement experiment (Prolific, LimeSurvey, R;
N = 79: N_{uk} = 40, N_{us} = 39; F = 41, M = 38).

4 × 9 = 36 **experimental items** of the form:

- V [$_{CP}$ *that* NP$_i$...] CP
- V [[$_{PP}$ P NP$_i$] *and* [$_{CP}$ *that* Pron$_i$...]] PP & CP
- V [$_{PP}$ P [NP$_i$ *and* [$_{NP}$ *the fact* [$_{CP}$ *that* Pron$_i$...]]]] PP

For example:

- **CP** condition:
  Justin Bieber can depend [$_{CP}$ *that* [$_{NP}$ *his fans*]$_i$ *still love his early songs*].
- **PP & CP** condition:
  Justin Bieber can depend [[$_{PP}$ *on* [$_{NP}$ *his fans*]$_i$] *and* [$_{CP}$ *that they$_i$ still love his early songs*]].
- **PP** condition:
  Justin Bieber can depend [$_{PP}$ *on* [[$_{NP}$ *his fans*]$_i$ *and* [$_{NP}$ *the fact* [$_{CP}$ *that they$_i$ still love his early songs*]]]].

Thermometer method and **standard items**, as before.

Experimental Evidence 4

Another experiment (*familiar, depend, speak, talk*).

Acceptability judgement experiment (Prolific, LimeSurvey, R; N = 79: $N_{uk} = 40$, $N_{us} = 39$; F = 41, M = 38).

4 × 9 = 36 **experimental items** of the form:

- V [$_{CP}$ *that* NP$_i$...] — CP
- V [[$_{PP}$ P NP$_i$] *and* [$_{CP}$ *that* Pron$_i$...]] — PP & CP
- V [$_{PP}$ P [NP$_i$ *and* [$_{NP}$ *the fact* [$_{CP}$ *that* Pron$_i$...]]]] — PP

For example:

- CP condition:
  Justin Bieber can depend [$_{CP}$ *that* [$_{NP}$ *his fans*]$_i$ *still love his early songs*].
- PP & CP condition:
  Justin Bieber can depend [[$_{PP}$ *on* [$_{NP}$ *his fans*]$_i$] *and* [$_{CP}$ *that they$_i$ still love his early songs*]].
- PP condition:
  Justin Bieber can depend [$_{PP}$ *on* [[$_{NP}$ *his fans*]$_i$ *and* [$_{NP}$ *the fact* [$_{CP}$ *that they$_i$ still love his early songs*]]]].

Thermometer method and **standard items**, as before.

Experimental Evidence 4

Another experiment (*familiar, depend, speak, talk*).

Acceptability judgement experiment (Prolific, LimeSurvey, R;
N = 79: N_{uk} = 40, N_{us} = 39; F = 41, M = 38).

4 × 9 = 36 **experimental items** of the form:

- V [$_{CP}$ *that* NP_i...] — CP
- V [[$_{PP}$ P NP_i] *and* [$_{CP}$ *that* $Pron_i$...]] — PP & CP
- V [$_{PP}$ P [NP_i *and* [$_{NP}$ *the fact* [$_{CP}$ *that* $Pron_i$...]]]] — PP

For example:

- **CP** condition:
  Justin Bieber can depend [$_{CP}$ *that* [$_{NP}$ *his fans*]$_i$ *still love his early songs*].
- **PP & CP** condition:
  Justin Bieber can depend [[$_{PP}$ *on* [$_{NP}$ *his fans*]$_i$] *and* [$_{CP}$ *that they*$_i$ *still love his early songs*]].
- **PP** condition:
  Justin Bieber can depend [$_{PP}$ *on* [[$_{NP}$ *his fans*]$_i$ *and* [$_{NP}$ *the fact* [$_{CP}$ *that they*$_i$ *still love his early songs*]]]].

Thermometer method and **standard items**, as before.

Experimental Evidence 4

Another experiment (*familiar, depend, speak, talk*).

Acceptability judgement experiment (Prolific, LimeSurvey, R;
N = 79: N_{uk} = 40, N_{us} = 39; F = 41, M = 38).

4 × 9 = 36 **experimental items** of the form:

- V [$_{CP}$ *that* NP$_i$...] — CP
- V [[$_{PP}$ P NP$_i$] *and* [$_{CP}$ *that* Pron$_i$...]] — PP & CP
- V [$_{PP}$ P [NP$_i$ *and* [$_{NP}$ *the fact* [$_{CP}$ *that* Pron$_i$...]]]] — PP

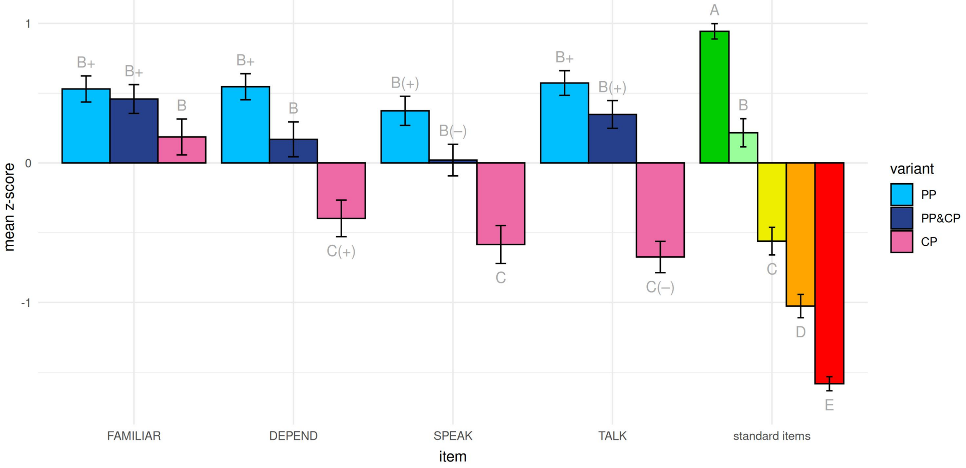
For example:

- **CP** condition:
  Justin Bieber can depend [$_{CP}$ *that* [$_{NP}$ *his fans*]$_i$ *still love his early songs*].
- **PP & CP** condition:
  Justin Bieber can depend [[$_{PP}$ *on* [$_{NP}$ *his fans*]$_i$] *and* [$_{CP}$ *that they*$_i$ *still love his early songs*]].
- **PP** condition:
  Justin Bieber can depend [$_{PP}$ *on* [[$_{NP}$ *his fans*]$_i$ *and* [$_{NP}$ *the fact* [$_{CP}$ *that they*$_i$ *still love his early songs*]]]].

Thermometer method and **standard items**, as before.

Experimental Evidence 5

Average z-scores (with 95% confidence intervals):



Experimental Evidence 6

Results:

- **for all predicates**, the acceptability of PP & CP is lower than that of PP (and higher than that of CP);
- for 1 predicate (*familiar*): **no statistically significant difference between PP & CP and PP**, **but CP (also) grammatical**;
- other 3 (*depend, speak, talk*): **statistically significant difference**...
  - **unexpected on asymmetric theories of coordination**,
  - **expected on symmetric theories of coordination**,
- ...and **CPs again at mid-range acceptability**,
  - suggesting again **gradient grammaticality**.

Experimental Evidence 6

Results:

- **for all predicates**, the **acceptability of PP & CP is lower than that of PP** (and higher than that of CP);
- for 1 predicate (*familiar*): **no statistically significant difference between PP & CP and PP**, **but CP (also) grammatical**;
- other 3 (*depend, speak, talk*): **statistically significant difference...**
  - **unexpected on asymmetric theories of coordination**,
  - **expected on symmetric theories of coordination**,
- ...and **CPs again at mid-range acceptability**,
  - suggesting again **gradient grammaticality**.

Experimental Evidence 6

Results:

- **for all predicates**, the **acceptability of PP & CP is lower than that of PP** (and higher than that of CP);
- for 1 predicate (*familiar*): **no statistically significant difference between PP & CP and PP**, **but CP (also) grammatical**;
- other 3 (*depend*, *speak*, *talk*): **statistically significant difference**...
  - **unexpected on asymmetric theories of coordination**,
  - **expected on symmetric theories of coordination**,
- ...and **CPs again at mid-range acceptability**,
  - suggesting again **gradient grammaticality**.

Experimental Evidence 6

Results:

- **for all predicates**, the **acceptability of PP & CP is lower than that of PP** (and higher than that of CP);
- for 1 predicate (*familiar*): **no statistically significant difference between PP & CP and PP**, **but CP (also) grammatical**;
- other 3 (*depend*, *speak*, *talk*): **statistically significant difference**...
  - **unexpected on asymmetric theories of coordination**,
  - **expected on symmetric theories of coordination**,
- ...and **CPs again at mid-range acceptability**,
  - suggesting again **gradient grammaticality**.

Experimental Evidence 6

Results:

- **for all predicates**, the **acceptability of PP & CP is lower than that of PP** (and higher than that of CP);
- for 1 predicate (*familiar*): **no statistically significant difference between PP & CP and PP**, **but CP (also) grammatical**;
- other 3 (*depend*, *speak*, *talk*): **statistically significant difference**...
  - **unexpected on asymmetric theories of coordination**,
  - **expected on symmetric theories of coordination**,
- ...and **CPs again at mid-range acceptability**,
  - suggesting again **gradient grammaticality**.

Experimental Evidence 6

Results:

- **for all predicates**, the **acceptability of PP & CP is lower than that of PP** (and higher than that of CP);
- for 1 predicate (*familiar*): **no statistically significant difference between PP & CP and PP**, **but CP (also) grammatical**;
- other 3 (*depend*, *speak*, *talk*): **statistically significant difference**...
  - **unexpected on asymmetric theories of coordination**,
  - **expected on symmetric theories of coordination**,
- ...and **CPs again at mid-range acceptability**,
  - suggesting again **gradient grammaticality**.

Experimental Evidence 6

Results:

- **for all predicates**, the **acceptability of PP & CP is lower than that of PP** (and higher than that of CP);
- for 1 predicate (*familiar*): **no statistically significant difference between PP & CP and PP**, **but CP (also) grammatical**;
- other 3 (*depend*, *speak*, *talk*): **statistically significant difference**...
  - **unexpected on asymmetric theories of coordination**,
  - **expected on symmetric theories of coordination**,
- ...and **CPs again at mid-range acceptability**,
  - suggesting again **gradient grammaticality**.