

Negative Polarity Items

A Semantic Property That Predicts a Syntactic Distribution

James Pustejovsky

Brandeis University
Department of Computer Science

CS 135: Computational Semantics

Fall 2026 (Session 5B — supplement to Session 5A)

The Phenomenon

Some expressions are **picky about where they occur**.

OK *Alice didn't see anyone.*

* *Alice saw anyone.*

OK *Nobody has ever been to Mars.*

* *Somebody has ever been to Mars.*

OK *Bob didn't lift a finger to help.*

* *Bob lifted a finger to help.* (on the idiomatic reading)

These are **negative polarity items** (NPIs). The obvious first hypothesis — *they need a negation* — is close, and wrong.

Today: what actually licenses them, why it is a monotonicity fact, and how far the generalization holds.

The Inventory

English has a large and heterogeneous NPI vocabulary.

Kind	Examples	Note
Determiners	<i>any, a single</i>	
Adverbs	<i>ever, yet, at all, anymore</i>	
Minimizers	<i>a red cent, a wink, a drop</i>	literal small quantities
Idioms	<i>lift a finger, budge an inch</i>	idiomatic reading only
Durative	<i>in years, in weeks</i>	stricter licensing
Aspectual	<i>until (punctual), either</i>	stricter licensing

Note the last two rows. Not all NPIs are licensed in the same environments — some need a stronger licenser. That is the *strength hierarchy*, and we return to it.

Minimizers are revealing: *he didn't drink a drop* means he drank nothing at all. The literal smallness is doing semantic work.

The Mirror Image: Positive Polarity

Some expressions are picky in the *opposite* direction.

OK Alice saw someone.

* Alice didn't see someone. (narrow-scope reading)

OK Bob would rather stay.

* Bob wouldn't rather stay.

Positive polarity items (PPIs) are *anti-licensed* in the very environments that license NPIs.

Alice didn't see someone is not ungrammatical — it has a **wide-scope** reading: $\exists x \neg \text{see}(a, x)$, “there is someone she didn't see”. What is blocked is the narrow-scope reading, where *someone* sits inside the negation.

So NPIs and PPIs partition environments between them — which is already evidence that a single semantic property is doing the work.

The Licensing Data

Environment	Example	
sentential negation	<i>Alice didn't <u>ever</u> call</i>	OK
<i>no</i> (scope)	<i>No student had <u>any</u> money</i>	OK
<i>every</i> (restrictor)	<i>Every student who had <u>any</u> money left</i>	OK
<i>every</i> (scope)	<i>Every student had <u>any</u> money</i>	*
<i>some</i> (scope)	<i>Some student had <u>any</u> money</i>	*
<i>few</i>	<i>Few students said <u>anything</u></i>	OK
<i>a few</i>	<i>A few students said <u>anything</u></i>	*
conditional antecedent	<i>If you have <u>any</u> questions, ask</i>	OK
conditional consequent	<i>If you ask, I have <u>any</u> answers</i>	*
<i>before</i>	<i>She left <u>before anyone</u> arrived</i>	OK
<i>after</i>	<i>She left <u>after anyone</u> arrived</i>	*
<i>without</i>	<i>He left <u>without any</u> warning</i>	OK
comparative	<i>Taller than <u>anyone</u> here</i>	OK

Thirteen environments. What do the seven acceptable ones share? Not negation — only two of them contain a negative word.

Ladusaw's Generalization

Ladusaw (1979)

An NPI is licensed exactly in a **downward-entailing** position.

Environment	Polarity	NPI
sentential negation	—	OK
<i>no</i> (scope)	—	OK
<i>every</i> (restrictor)	—	OK
<i>every</i> (scope)	+	*
<i>some</i> (scope)	+	*
<i>few</i>	—	OK
<i>a few</i>	+	*
conditional antecedent	—	OK
<i>before</i>	—	OK
<i>after</i>	+	*

A perfect match. A semantic property — behaviour under set inclusion — predicts a syntactic distribution, with no syntactic stipulation anywhere.

The Crucial Case

OK *Every student who had any money left*

* *Every student had any money*

Same determiner. Same NPI. Different argument.

Recall that *every* is $\downarrow\text{MON}\uparrow$ — downward in its restrictor, upward in its scope. The NPI is licensed in the first and only the first.

No theory that treats *every* as uniformly “positive” or “negative” can get this right.

The licensing condition has to be stated *per argument position*, which is exactly what monotonicity does and what a lexical feature like $[\pm\text{NEG}]$ cannot.

This single contrast is the strongest evidence for Ladusaw’s generalization, and it is why we spent Session 5A separating left from right monotonicity.

How to Test a Position for DE

A procedure you can apply to any environment.

- 1 Find a pair of predicates where one is more specific: $\text{jog} \subseteq \text{move}$, $\text{poodle} \subseteq \text{dog}$.
- 2 Put the **general** one in the position and check the sentence is true.
- 3 Substitute the **specific** one.
- 4 If truth is preserved, the position is downward entailing.

Worked: the before-clause.

She left before anyone arrived

Test: *She left before a student arrived* — does *She left before a first-year student arrived* follow?

Yes. If she left before any arrival at all, she left before the more specific ones too. **DE** ✓

Now try *after*: *She left after a student arrived* does *not* entail *She left after a first-year student arrived*. Not DE — and correspondingly it blocks NPIs.

Not All Downward Contexts Are Equal

Ladusaw's generalization is not quite the whole story. Consider:

OK *At most three students have ever been to Paris.*

? *At most three students have been to Paris in years.*

OK *No student has ever been to Paris.*

OK *No student has been to Paris in years.*

Both *at most three* and *no* are downward entailing. But only *no* licenses *in years*.

So NPIs come in strengths, and DE-ness comes in strengths to match. The rest of this session is about that correspondence.

Ever and *any* are **weak** NPIs; *in years*, *until*, and *either* are **strong**.

Anti-Additivity

A strictly stronger property than downward entailment.

Anti-additive

f is anti-additive iff $f(A \vee B) \equiv f(A) \wedge f(B)$

Test it on no:

No student sang or danced $\equiv$ *No student sang **and** no student danced* ✓

Test it on at most three:

At most three students sang or danced

Not equivalent to *at most three sang and at most three danced*. Suppose two sang and two others danced: both conjuncts hold, but four students sang-or-danced. *

So *no* is anti-additive and *at most three* is merely DE — exactly tracking which one licenses *in years*.

The Zwarts Hierarchy

Three nested classes of negative strength, with NPI classes to match.

Class	Condition	Licensors	Licenses
Downward entailing	$A \subseteq B \Rightarrow f(B) \subseteq f(A)$	<i>few, at most three</i>	weak: <i>any, ever</i>
Anti-additive	$+ f(A \vee B) = f(A) \wedge f(B)$	<i>no, without</i>	strong: <i>in years</i>
Antimorphic	$+ f(A \wedge B) = f(A) \vee f(B)$	<i>not (sentential)</i>	superstrong

antimorphic $\subsetneq$ anti-additive $\subsetneq$ downward entailing

A graded licensing condition. The stronger the NPI, the stronger the licenser it demands — and the classes line up exactly.

Structure Matters, Not Word Order

Linear precedence is not enough. The licenser must be **structurally superior** to the NPI.

OK *No one who has any money is here.*

* *Someone who has no money has any friends.*

In the second, *no* precedes *any* linearly — but it is buried inside a relative clause modifying the subject, and does not take scope over the main predicate.

The licenser must c-command the NPI — equivalently, the NPI must lie inside the licenser's semantic scope. Which is just to say: polarity is computed over the *derivation*, not over the string.

This matters computationally. A model that keys on the presence of a negative word nearby will get these wrong, and the minimal pairs to catch it are easy to construct.

Cases That Strain the Generalization

1. **Only.** Licenses NPIs, but is not DE in the relevant argument.

Only Alice had any money. OK

2. **Questions.** License NPIs with no DE operator in sight.

Did anyone call? OK

3. **Comparatives.**

Alice is taller than anyone here. OK

4. **The restrictor of every** is DE only if the conclusion's presupposition is met.

Four genuine counterexamples. The generalization is too good to abandon and too leaky to leave as it stands — so the literature repaired it.

Von Fintel's Repair: Strawson-DE

Strawson downward entailment (von Fintel 1999)

A position is Strawson-DE if the inference goes through **whenever the conclusion's presuppositions are satisfied**.

Apply it to only.

Only Alice sings. $\Rightarrow$ *Only Alice sings loudly.*

This fails in general — but *only Alice sings loudly* presupposes that Alice sings loudly. Grant that presupposition, and the inference goes through.

So only is Strawson-DE, and Ladusaw's generalization survives — restated with Strawson-DE in place of plain DE.

Worth noticing the shape of this episode: a clean generalization, real counterexamples, and a principled weakening that keeps the predictive content. That is what progress looks like here.

Why Would Language Work This Way?

Ladusaw's generalization describes. It does not explain.

The domain-widening account (Kadmon & Landman 1993; Krifka; Chierchia).

An NPI like *any* **widens the domain of quantification** — it signals that even marginal cases are included.

Context	Widening makes the claim	Result
downward	stronger	informative, licensed
upward	weaker	gratuitously weak, blocked

Check it. *I don't have any money* is stronger than *I don't have money* — it rules out even a penny. But *I have any money* would be weaker than *I have money*, and asserting a deliberately weakened claim is odd.

So the *distribution* is predicted by monotonicity, and the *reason* is pragmatic. A nice division of labour.

NPIs as a Probe for Neural Models

NPI licensing is an unusually good diagnostic, and it is used as one.

- **Minimal pairs are easy to build automatically**, differing in one word.
- **Acceptability is sharp**, and native judgements agree.
- **Getting it right requires hierarchical structure**, not surface cues — the c-command slide.
- **The strength hierarchy gives graded difficulty** for free.

BLiMP and the NPI-licensing probes of Warstadt et al. are built on exactly this. Models do well on the easy cases and less well where the licenser is structurally embedded — consistent with keying on surface proximity.

And notice what building such a benchmark requires: the DE classification, the strength hierarchy, and the c-command condition. You cannot test a model for this without first having the theory.

Summary

Point	Detail
NPIs	<i>any, ever, lift a finger, in years, ...</i>
PPIs	the mirror image: <i>some, would rather</i>
Ladusaw	licensed exactly in downward-entailing positions
The crucial case	<i>every</i> licenses in its restrictor, not its scope
Testing DE	substitute a more specific predicate; check truth survives
Strength	weak (DE), strong (anti-additive), superstrong (antimorphic)
Anti-additive	$f(A \vee B) = f(A) \wedge f(B)$; <i>no yes, at most three no</i>
c-command	structure, not linear order
Strawson-DE	von Fintel's repair; rescues <i>only</i>
Domain widening	why: widening strengthens in DE, weakens in UE
As a probe	sharp judgements, automatic minimal pairs, needs structure

Reading and Exercises

Reading:

- Ladusaw (1979), *Polarity Sensitivity as Inherent Scope Relations*
- Zwarts (1998), *Three Types of Polarity*
- von Stechow (1999), *NPI Licensing, Strawson Entailment, and Context Dependency*
- Chierchia (2013), *Logic in Grammar* — the pragmatic derivation
- Warstadt et al. (2019) on NPI licensing as a probe

Exercises:

- 1 Classify *seldom*, *rarely*, *hardly* for DE, anti-additivity, antimorphism.
- 2 Construct a minimal pair distinguishing linear precedence from c-command.
- 3 *Nobody who didn't bring any money was admitted* — is *any* licensed? Compute the polarity and check against your judgement.
- 4 Find an environment where your judgement and Ladusaw's generalization disagree. Then decide which one is wrong.