

---

# Towards a semantics for letter predicates\*

Mathieu Paillé • U. of Calgary

CLA • McGill • June 4, 2025

---

## 1 Introduction

- A class of predicates that has received almost no attention: **symbol predicates**.<sup>1</sup> These are predicates whose meanings are symbols, like letters or written numbers.
- I focus on letter predicates—words like ‘f’ /ɛf/. They apparently can be true of either graphemes or phonemes:

- (1) a. (Phonologically,) there are two ‘f’s in ‘philosophy.’  
b. (Orthographically,) there are two ‘f’s in ‘traffic.’

- To be sure, letter names can also be used as arguments (2); but we’ll focus on their predicative use today.

- (2) ‘B’ is the second letter of the alphabet.

- Section 2 focuses on (1) and suggests a first hypothesis about letter predicates’ meanings.
- Sections 3 then modifies the proposed meaning to integrate two layers of context-sensitivity in letter predicates’ meanings.
- Section 4 discusses ‘co-predications’ of letter predicates like (3), finding that the final hypothesis from section 3 can explain them (in an arguably counter-intuitive way).

- (3) In this language, the ‘d’s are ‘g’s.

---

\*I thank Nina Haslinger, Bernhard Schwarz, Benjamin Spector, the audience at the 2024 Alberta Conference on Linguistics (ACOL) at the University of Calgary, and of course the CLA reviewers.

<sup>1</sup>I only know of Gasparri 2019 on written numbers.

## 2 An initial hypothesis for letter predicates

- This section is an initial foray into letter predicates, focusing on data like (1) or (4).

- (4) There is one ‘f’ in this word.  
a.  $\rightsquigarrow$  one  $\langle f \rangle$   
b.  $\rightsquigarrow$  one  $/f/$

### 2.1 Letters as phonemes

- We’ve seen that letter predicates can sometimes be used to refer to a phoneme (5a). In fact, letter predicates can be applied to individuals explicitly stated to be sounds (5b):

- (5) a. There are two ‘f’s in ‘philosophy.’  
b. That sound is an ‘f.’

- I assume sounds cannot have visual forms, so (5b) cannot mean that the sound has the form  $\langle f \rangle$ .

- Let’s start building up a lexical meaning for letter predicates:

- (6)  $\llbracket \text{‘f’} \rrbracket = \lambda x. \text{occurrence-of}(x, /f/)$ .

- (6) uses the meta-language predicate **occurrence**:

- An ‘f’ is not the TYPE  $/f/$ , nor is it necessarily just a TOKEN of  $/f/$ —tokens are spatiotemporally concrete entities (Wetzel 2018).
- Many ‘f’s are not spatiotemporally concrete, as in (5a) (a statement about the word *philosophy*, an abstract entity).
- So we say there are two ‘occurrences’ of  $/f/$  in *philosophy*.

- For letters like ‘a’ associated with multiple phonemes, we’d need disjunctive meanings:

- (7)  $\llbracket \text{‘a’} \rrbracket = \lambda x. \text{occurrence-of}(x, /æ/) \vee \text{occurrence-of}(x, /ɑ/) \vee \text{occurrence-of}(x, /eɪ/)$ .

- The assumption here is that English speakers’ lexical meanings for letter predicates is based on English phonemes. We return to this in section 3.

## 2.2 Letters as graphemes

- (6)–(7), however, constrain the meaning of letter predicates too much—we already saw data like:
- (8) There are two ‘f’s in ‘traffic.’
- $\approx$  two occurrences of  $\langle f \rangle$  (true meaning)
  - $\not\approx$  two occurrences of  $/f/$  (possible but false meaning)
- Other examples where it seems we are talking about graphemes, not phonemes:
- (9) a. There are two ‘p’s in ‘philosophy.’  
b. There is a ‘q’ in ‘qi.’  
c. There is an ‘l’ in ‘salmon.’
- As a first attempt, we could posit lexical ambiguity between the PHONEME-meaning and the GRAPHEME-meaning of letter predicates.
- (10)  $\llbracket \text{‘f’} \rrbracket =$
- $\lambda x. \text{occurrence-of}(x, \langle f \rangle).$  ‘GRAPHEME meaning’
  - $\lambda x. \text{occurrence-of}(x, /f/).$  ‘PHONEME meaning’
- However, lexical ambiguity might not be quite right, due to data like:
- (11) a. SCENARIO: *As part of a modern art exhibit, an artist sets up a room where there is nothing but a canvas with a big ‘f’ painted on it, and a speaker continuously playing a recorded /f/.*  
b. I dislike both ‘f’s in this room.
- If there was lexical ambiguity, (11b) would be infelicitous in that scenario.
  - ‘f’ could only be interpreted on the GRAPHEME meaning or the PHONEME meaning; either way, there is just one ‘f.’
  - Instead of ambiguity, I therefore posit a disjunctive meaning for letter terms:
- (12)  $\llbracket \text{‘f’} \rrbracket = \lambda x. \text{occurrence-of}(x, \langle f \rangle) \vee \text{occurrence-of}(x, /f/).$
- An issue at the moment is how to constrain (12) so that in a particular sentence, a speaker can use a letter predicate to refer *exclusively* to graphemes or phonemes. We return to this in section 3.2.

## 3 Two layers of context-sensitivity

### 3.1 Context-sensitivity 1: writing-system relations

- Even with (12), lexicalizing reference to phonemes is still a questionable move.
  - Consider letters’ variation across languages. Let’s start with (13), which does not appear problematic: it can be captured through the GRAPHEME-meaning disjunct of (12).
- (13) In French, ‘u’s are pronounced as  $/y/$ .  
 $\approx$  ‘In French, occurrences of the grapheme  $\langle u \rangle$  are given the pronunciation  $/y/$ .’
- But things are harder with (14). This example relies on the fact that the Arabic letter ghayn  $\langle \text{غ} \rangle$  is pronounced  $/ʁ/$ , like  $\langle r \rangle$  in French.
- (14) Ghayn (in Arabic) is a French ‘r.’
- So far, in modelling English speakers’ meanings for letter terms, I have been assuming reference to *English* phonemes, so that ‘r’ means:
- (15)  $\llbracket \text{‘r’} \rrbracket = \lambda x. \text{occurrence-of}(x, \langle r \rangle) \vee \text{occurrence-of}(x, /ɹ/).$
- Yet, (14) equates ghayn with neither of the disjuncts in (15). What we need is reference to the phoneme  $/ʁ/$ .
  - Of course, part of the story is that ‘r’ in (14) is modified by *French*. But with the lexical meaning in (15), it’s not clear how ‘r’ and *French* even compose.
    - If anything, one would expect *French* ‘r’ to necessarily pick out the GRAPHEME-meaning, since there are occurrences of  $\langle r \rangle$  in French but not of  $/ɹ/$ .
  - We presumably don’t want to lexicalize reference to every phoneme a letter can be associated with across all languages.
  - So I’ll take it that knowledge about which graphemes correspond to which phonemes is fed semantically into letter predicates in the form of a contextually provided relation from graphemes to phonemes.

- Assume speakers are familiar with one or more **WRITING CONVENTIONS**—anything from e.g. Italian writing conventions (16) to more specific things like the convention for Mandarin borrowings in English.<sup>2</sup>
- The **DOMAIN**  $D_R$  of the writing-convention relation  $R$  is the set of **GRAPHEMES** in that writing system. The **RANGE**  $R_R$  is the set of **PHONEMES**.

#### Italian writing convention

$$(16) \quad \begin{array}{lcl} \langle a \rangle & \longrightarrow & /a/ \\ \langle b \rangle & \longrightarrow & /b/ \\ \langle g \rangle & \begin{array}{l} \longrightarrow \\ \searrow \end{array} & \begin{array}{l} /g/ \\ /dʒ/ \end{array} \\ \dots & & \end{array}$$

- Lexically, then, letters take a writing convention as their first argument:

$$(17) \quad \begin{aligned} \llbracket 'f' \rrbracket &= \lambda R : \mathbf{writing-convention}(R). \lambda x. \\ &\exists y \in D_R [y = \langle f \rangle \wedge \mathbf{occurrence-of}(x, y)] \vee \\ &\exists y \exists z [y = \langle f \rangle \wedge y R z \wedge \mathbf{occurrence-of}(x, z)] \end{aligned}$$

- (17) adds context-sensitivity in letter predicates' meanings in letting them be used vis-à-vis a particular writing convention.
- The first disjunct in (17) states the same thing as the first disjunct in our original (12), repeated in (18). The difference is the second disjunct: (18) lexicalizes reference to particular phonemes, while (17) lets a writing system argument provide the phonemic value.

$$(18) \quad \llbracket 'f' \rrbracket = \lambda x. \mathbf{occurrence-of}(x, \langle f \rangle) \vee \mathbf{occurrence-of}(x, /f/).$$

- The meaning in (17) is **asymmetric**; it centres the graphemic component of letter predicates, with the phoneme parasitic on that grapheme.

### 3.2 Context-sensitivity 2: Phoneme-only, grapheme-only uses of letter predicates

- As we've seen, another kind of context-sensitivity is that letter predicates can be used exclusively for phonemes (19a), for graphemes (19b), or for both (19c):

- $$(19) \quad \begin{array}{ll} \text{a.} & \text{There is one 'f' in 'traffic.'} \\ \text{b.} & \text{There are two 'f's in 'traffic.'} \\ \text{c.} & \text{I dislike both 'f's in this room. (in the art-installation scenario)} \end{array}$$

- While (17) captures (19c), it cannot capture (19a–b), where the speaker is only counting one 'kind' of 'f': 'f'-graphemes or 'f'-phonemes.
- (20) adds this extra piece of context-sensitivity—the  $S$  variable (a set) lets speakers limit the extension of 'f' to graphemes, phonemes, or both.<sup>3</sup>

$$(20) \quad \begin{aligned} \llbracket 'f' \rrbracket &= \lambda R : \mathbf{writing-convention}(R). \lambda S : S = D_R \vee S = R_R \vee S = D_R \cup R_R. \lambda x. \\ &\exists y [y \in S \wedge y = \langle f \rangle \wedge \mathbf{occurrence-of}(x, y)] \vee \\ &\exists y \exists z [z \in S \wedge y = \langle f \rangle \wedge y R z \wedge \mathbf{occurrence-of}(x, z)]. \end{aligned}$$

- (20) is the final hypothesis for today.

## 4 Letter co-predications

- I conclude by turning to a new class of data: co-predications—the application of two letter predicates to a single (plural or singular) individual.

- We already saw one such example:

$$(21) \quad \text{An Arabic ghayn is a French 'r.'}$$

- The meaning of (21) is that the grapheme  $\langle \text{ġ} \rangle$  represents  $/r/$ .

- Stepping back from this particular example, we observe in letter-term co-predications not just the 'pronounced as' meaning of (21), but also a 'written as' meaning:

$$(22) \quad \text{In this language, the 'd's are 'g's.}$$

- $$\begin{array}{ll} \text{a.} & \rightsquigarrow \langle d \rangle \text{ pronounced } /g/ \\ \text{b.} & \rightsquigarrow /d/ \text{ written } \langle g \rangle \end{array}$$

- To reiterate from section 2: I assume that (an occurrence of) a grapheme can't 'be' (an occurrence of) a phoneme, or vice-versa. So can we understand (22)?

<sup>3</sup>A problem at the moment, which I will not fix today, is how to rule out (i):

(i) #There are three 'f's in 'traffic.'

After all, there are two orthographic 'f's and one phonological 'f.' See Liebesman & Magidor 2025 for extensive discussion of similar problems.

<sup>2</sup>Writing conventions are obviously more complicated than shown here; they presumably include information about e.g. digraphs ( $\langle ph \rangle$  as  $/f/$ ), context-sensitive spelling rules, exceptions, etc.

#### 4.1 The ‘pronounced as’ meaning—(21) and (22a)

- The truth conditions of (22a)/(23) can be captured with a slightly counter-intuitive move: *the ‘d’s* is interpreted on the PHONEME, not GRAPHEME, disjunct.

(23) In this language, the ‘d’s are ‘g’s.

a.  $\rightsquigarrow \langle d \rangle$  pronounced  $/g/$

- Let  $R_e$  be the writing convention for English and  $R_l$  be the writing convention for whatever language (23) refers to. For ‘g’s, we feed ‘g’ the English writing convention and a set  $S = R_{R_e}$  (i.e. English phonemes).

(24)  $\llbracket \llbracket \text{‘g’ } R_e \rrbracket R_{R_e} \rrbracket = \lambda x. \exists y \exists z [z \in R_{R_e} \wedge y = \langle g \rangle \wedge y R z \wedge \text{occurrence-of}(x, z)]$ .  
 ( $\approx x$  is an occurrence of a/the phonemic value of  $\langle g \rangle$  in English)<sup>4</sup>

- We use the same PHONEME meaning for ‘d’ in *the ‘d’s*:

(25)  $\llbracket \llbracket \text{‘d’ } R_l \rrbracket R_{R_l} \rrbracket = \lambda x. \exists y \exists z [z \in R_{R_l} \wedge y = \langle d \rangle \wedge y R z \wedge \text{occurrence-of}(x, z)]$ .

- Putting aside how exactly (24) and (25) compose,<sup>5</sup> here’s the intuitive paraphrase: “the phoneme of ‘d’ in that language = the phoneme of ‘g’ in English.”
  - Even though it really feels like we are talking about the grapheme  $\langle d \rangle$  (and asserting it represents  $/g/$ ), the truth conditions are actually captured through the PHONEME meaning of both letter predicates.

#### 4.2 The ‘written as’ meaning—(22b)

- The ‘written as’ meaning of (22b)/(26) is just the mirror image of the ‘pronounced as’ meaning.

(26) In this language, the ‘d’s are ‘g’s.

b.  $\rightsquigarrow /d/$  written  $\langle g \rangle$

- The intuitive explanation: (26) equates the phoneme of English  $\langle d \rangle$  with the phoneme of the  $\langle g \rangle$  of that language—hence meaning that  $\langle g \rangle$  is pronounced  $/d/$ .
- Concretely, this means that the only difference from the ‘pronounced as’ meaning is that we switch which letter predicate is fed  $R_e/R_{R_e}$  and which one is fed  $R_l/R_{R_l}$ :

(27) a.  $\llbracket \llbracket \text{‘d’ } R_e \rrbracket R_{R_e} \rrbracket = \lambda x. \exists y \exists z [z \in R_{R_e} \wedge y = \langle d \rangle \wedge y R z \wedge \text{occurrence-of}(x, z)]$ .  
 b.  $\llbracket \llbracket \text{‘g’ } R_l \rrbracket R_{R_l} \rrbracket = \lambda x. \exists y \exists z [z \in R_{R_l} \wedge y = \langle g \rangle \wedge y R z \wedge \text{occurrence-of}(x, z)]$ .

## 5 Conclusion

- Letter predicates are surprisingly interesting.
- They are context-dependent in at least two ways, taking both a writing convention and a variable corresponding to graphemes vs. phonemes vs. both.
- There are a number of remaining questions and interesting data, including modelling the relationship between letters as types (‘A is a letter’) and letters as occurrences (‘This is an A’).

## References

- Gasparri, Luca. 2019. A numeral oddity. *Journal of Semantics* 36:563–571.
- Liebman, David, & Ofra Magidor. 2025. *Property versatility and copredication*. Oxford: Oxford University Press.
- Wetzel, Linda. 2018. Types and Tokens. In *The Stanford encyclopedia of philosophy*, ed. Edward N. Zalta. Metaphysics Research Lab, Stanford University, Fall 2018 edition.

<sup>4</sup>I suppressed the left disjunct—the GRAPHEME meaning—for simplicity of presentation.

<sup>5</sup>The details aren’t fully clear to me yet, but presumably we access the individual concept for *the ‘d’s*, and say through an equative *be* that their phonemic value in this world is the same as the phonemic value of English ‘g.’