

Distributional profiling and the semantics of modifier classes

Martin Schäfer

Abstract The versatility of modifiers makes their analysis a challenge. Their interpretation depends on their own lexical semantics, their syntactic position, and the semantics of the material they combine with. This paper explores what distributional profiling reveals about four modifiers that are traditionally taken to correspond to different semantic classes. The profiles of four adjectives are presented and the results are qualitatively interpreted. While the results show that assumptions made in the literature are partly reflected in the distributional data, they also show many patterns that require further explanation.

Keywords modification · distributional semantics · English · adjective · adverb · adverbial

M. Schäfer, Eberhard Karls Universität Tübingen, <https://www.martinschaefer.info/>

In Christopher Pinon & Laurent Roussarie (eds.), *Empirical Issues in Syntax and Semantics 13*, 139–166. Paris: CSSP. <http://www.cssp.cnrs.fr/eiss13/>
© 2020 Martin Schäfer

1 Introduction

Pairs like English *quick/quickly*, *slow/slowly*, *wise/wisely*, and *lucky/luckily* can occur in a variety of positions. This syntactic versatility alone makes fully understanding their behavior a challenge. On top of this, the lexical semantics interacts with their combinatorial potential in intricate ways. Even within semantic classes there is considerable variation in behavior, raising the question of the extent to which investigations of individual items in individual constructions yield usable generalizations. This paper explores the possibility of using distributional semantics across the usages of these items in order to come to interpretable patterns of the behavior of different adjective/adverb pairs.

The pairs were selected because they fulfill the following syntactic and semantic criteria: in their base form, they occur in attributive as well as predicative position. In predicative position, they allow the combination with *to*-INF (initial). In their *-ly* form, they occur in sentence-initial, pre-

verbal and postverbal position, and exhibit a number of different, position-dependent, readings. Further, the occurrence in the *to*-INF construction has been described as expressing the same meaning as their occurrence as *-ly* adverbs.

This paper focuses on these patterns, three adjectival and three adverbial patterns, for each of the pairs. They are illustrated for *wise/ly* in (1) and (2), with examples from the British National Corpus (BNC).

- (1) a. attributive adjective [attrib]
 “A most wise precaution,” Karl said. [A7A 3043]
- b. *to*-infinitival [INF]
 Perhaps Mrs Nicholson had been wise to leave. [AT4 892]
- c. predicative [pred]
 If that is so, he is wise. [AKY 830]
- (2) a. sentence-initial adverb [advSI]
 Wisely, Bright has included biographical entries of dead linguists only. [J7K 33]
- b. preverbal adverb [advV]
 The CO wisely decided not to notice this particular instance of it.
 [ACE 2163]
- c. postverbal adverb [Vadv]
 We help you choose wisely.
 [A65 1983]

Note that they are all treated as mutually exclusive, e.g. *predicative* translates to ‘predicative but not followed by *to*-INF’.

The paper is structured as follows: §2 presents a short overview of work on modifier classes as relevant to this investigation. §3 gives an overview of the occurrence frequencies for the four pairs in the respective pattern in the ukWaC, a 2 billion word corpus of English. §4 presents the distributional analysis. The results are discussed in §5, and §6 concludes.

2 Lexeme type, modifier classes, and available readings

One of the main motivations for the exploration of distributional profiles for these four adjectives is the difficulty of getting to grips with them with

other approaches. In particular, it is difficult to reliably establish classes without appealing to intuition, and even if one has established plausible classes, it is not easy to show that they are also linguistically and/or cognitively relevant.

At the level of lexical semantics, the four adjectives, *quick*, *slow*, *wise*, and *lucky*, have been discussed to various degrees. A predicate like *quick* is often seen as an event predicate, cf. Pustejovsky (1995) for *fast* and Bücking & Maienborn (2019) for several of its German translation equivalents. For these authors, the main question is how such an event predicate can successfully be combined with non-event predicates, e.g. artefacts (*fast car*, etc.). A similar judgement should hold for its antonym *slow*. Geuder (2002: 10) categorizes *slow* as an external property connected to movement/change, and thus expected to be “primarily a property of dynamic entities – i.e. events, not individuals”.

Geuder (2002) also discusses *wise* and *lucky*. Similar to his view (p. 10) on *intelligent* as connected to psychological conditions, he says that *wise* is a disposition of an individual. The disposition describes a capability (p. 113). *Wise* belongs to the group of agentive adjectives (p. 113), contrasting with evaluative adjectives, a class which contains *lucky*. On his analysis, neither are predicates of events. But there is a link to events for agentive adjectives (in his terms, they make covert reference to an event). If Geuder is right, what are the linguistic reflexes of these different adjective classes?

As evidenced by the availability of the four pairs in all six patterns under consideration (three for the adjectives and three for the related adverbs), there are no striking restrictions on their overall combinatorial possibilities (cf. attributive-only or never-attributive adjectives like *main* or *alone*). One example for a combinatorial constraint can be found in the standard predicative usage: only *wise* allows for an *of*-phrase indicating the agent (cf. Oshima 2009), as in (3).

(3) “That’s wise of you, miss.” [FR6 1043]

Of-phrases often occur together with *it*-extraposition, as in (4).

(4) ... , it would surely be wise of Althusser to show how he proposes to do it. [CMN 628]

Geuder (2002: 112), who also discusses *of*-phrases in distinguishing between agentive and evaluative adjectives, further points to a corresponding inability of agentive adjectives to occur with *for*-phrases, as in (5).

(5) That was clever of John/??for John. (= Geuder's (19))

However, at least for *wise*, the situation is not so clear-cut, as seen in examples like (6).

(6) Under those circumstances, I ask him whether it would be wise for the House to proceed with the Bill tonight. [HHX 17055]

In the BNC, the *wise for*-construction, with 30 occurrences, is distinctly more frequent than *wise of*, which occurs only ten times.

In their attributive usages, there is on the surface little difference between the four adjectives. For example, all four adjectives occur with heads referring to more than two different ontological types, all including at least events and physical entities, as in (7).

- (7)
- a. quick decision
 - b. quick antidote
 - c. slow progress
 - d. slow boat
 - e. lucky draw
 - f. lucky winner
 - g. wise counselling
 - h. wise dragon

Showing that there is a quantitative pattern behind this requires annotation of the ontological types of the heads. In Schäfer (2020a), I show that the majority of heads in the top 120 collocations for both *quick* and *slow* fall in the event category, which is in line with the assumption that these are event predicates.

Larson (1998: 18) and Bücking & Maienborn (2019: 35–36) remark that *quick* and *slow* are more restricted in the available readings in predicative position than in attributive position. Differences between *wise* and the other three adjectives should follow from the expectation that dispositions are better suited for attributive position than predicative position (cf.

Cruse 2004: 301), although I don't know any study showing this explicitly. Geuder (2002: 139–147) links fine-grained differences in the properties of clausal complements of his agentive and evaluative classes to his analysis of them as non-event predicates, and of the covert event reference of the former.

For the adverbial usages of the adjectives, the four adjectives show three patterns in alignment with their different lexical semantics. Of the four, *wisely* is the only one that participates in the standard high-low pattern, where a sentential reading contrasts with a manner reading (e.g. Ernst 2002). While the immediately preverbal position allows both readings, the sentence-initial position is linked to the high reading, and postverbal position to the low reading. In its high reading, *wisely* is a subject or rather agent-oriented sentence adverbial (cf. Maienborn & Schäfer 2011), and in its low reading, it is a manner adverbial. The semantic difference to manner usages is, among other things, revealed through paraphrases:

- (8) a. Wisely, Bright has included biographical entries of dead linguists only. [J7K 33]
 b. It was wise of Bright to have included biographical entries of dead linguists only.

In contrast, if manner usages are available, paraphrases like “in an ADJ manner/the way in which ... BE ADJ” are more apt:

- (9) a. The old lady nodded wisely: “I thought so ... scientists would have tried it out on rats first.” [A57 17]
 b. The old lady nodded in a wise manner/The way in which the old lady nodded was wise.

Quickly and *slowly*, aspect-manner adverbs in the terminology of Ernst (2002), also occur with different readings depending on position, as in (10).

- (10) a. Lynn quickly raced down the hallway. (= Ernst's (2.149a))
 b. Lynn raced down the hallway quickly. (= Ernst's (2.149b))

While the clarity of the readings depends on the specific verb, one can distinguish between inceptive, holistic, and true rate usages of *quickly*,

with inceptive and holistic readings aligned with high positions, and the true rate reading with the low position. An inceptive usage is one where *quickly* indicates that the time up until an event is short, a holistic usage one where the whole event referred to by the verb phrase took only a short amount of time, and a true rate usage one where the rate of an internal movement inherent to the event is targeted. While they are not always independent of each other (cf. §5.3), an inceptive reading of (10a) is one where *quickly* indicates that the racing-down-the-hallway event takes place shortly after some other, contextually supplied event. The holistic reading is one where the racing-down-the-hallway event itself takes only a short time. Finally, the true rate reading, associated with (10b), indicates that the running itself consists of quick movements, i.e. short subevents. Several terms have been used to describe these differences. Ernst (2002) refers to the inceptive readings as clausal readings of aspect-manner adverbs and takes them to be special cases of a manner reading. Cinque (1999) discusses the preverbal and postverbal occurrences of *quickly* in terms of two different aspectual projections, celerative aspect I and celerative aspect II. While the former involves quantification over an event, the latter involves quantification over a process. For Schäfer (2013) it's a matter of event-related vs. verb-related modification. The latter assumes that the modifier does not directly predicate over the event referred to by the verbal predicate but is connected to the event via some appropriate relation, as in the proposals regarding speed and manner in Dik (1975) and Piñón (2008), respectively.

Luckily, a sentential adverb from the evaluational subclass of speech-act adverbials (cf. Maienborn & Schäfer 2011), can be paraphrased as in (11b).

- (11) a. Luckily the flies had gone by now. [A0N 2400]
 b. It was lucky that the flies had gone by now.

According to Ernst (2002: 78), *luckily* is a pure evaluative and does not come with a manner reading (though it can occur as a verb modifier for some verbs, with a resultative-like interpretation, as in his (2.132a) *The performance turned out pretty luckily, considering the troubles we'd had*).

While these paraphrases already show that the same content can be

expressed by using either the adjective or the adverb of each pair, both *wisely* and *luckily* have been explicitly linked to the *to*-INF pattern. Oshima (2009: 364) points to the following two sentences with *wisely* as being “roughly synonymous” (an observation also made in earlier works):

- (12) a. Wisely, John left early. (= Oshima’s (1a))
 b. John was wise to leave early. (= Oshima’s (1b))

The corresponding two sentences with *luckily* show the same pattern:

- (13) a. Luckily, John passed the exam. (= Oshima’s (6a))
 b. John was lucky to pass the exam. (= Oshima’s (6b))

As shown in Schäfer (2020a), *to*-INF is equally frequent with *quick* and *slow*, and they also seem very close to their adverbial counterparts, as in (14).

- (14) a. Therefore, they are slow to respond to market changes as reflected by movements in relative prices. [HXL 133]
 b. Acne responds slowly and drugs need time to work. [CDR 1954]

For all four cases, the additional challenge is how to further tackle the differences within the spectrum of “rough synonymy”. Oshima (2009: 372–373), building on Wilkinson (1970) and Barker (2002), sees an assertion/presupposition reversal between the adverbial and the *to*-INF for both *wise* and *lucky*: e.g. that John left early is asserted in (15a) but presupposed in (15b).

- (15) a. Wisely, John left early.
 b. John was wise to leave early.

Karttunen (2013) argues that *lucky-to*-INF does not presuppose its complement, but is two-way implicative, that is, it yields a positive entailment in positive contexts, but a negative entailment in negative contexts. An example for the latter is (16), where the entailment is negative (“I did not get a table on this trip”) in the negative context provided by ordinary negation.

- (16) Anyway, I was not lucky to get a table on this trip. Maybe next time.
 (= Karttunen’s (3a))

In Schäfer (2020b), I point out a further difference between *lucky-to*-INF and *luckily*: whereas for the standard evaluative usage of *luckily* there is no restriction on the recipient of the luck (lucky for who?), *lucky-to*-INF is always restricted to a subject-oriented interpretation.

For *quick/slow-to*-INF and the adverbial usages, I (in Schäfer 2020a and Schäfer 2020b) try to establish a reliable difference. For both adjectives, there are minority usages with non-ordinary subjects patterning with the *tough*-construction, as in (17).

- (17) The following recipes are quick to prepare and very low in calories.
[CDR 220]

In the majority pattern, where *quick/slow-to*-INF take ordinary subjects, as in (18), *quick/slow-to*-INF share with *wise* the alignment with the high adverbial readings, that is, for these two, a true rate reading is excluded.

- (18) a. And he is quick to point out that it was a joint decision to make a serious bid. [G39 1207]
b. But foreign governments have been slow to respond with aid. [B7N 203]

Other than that, there are only tendencies, with *quick-to*-INF on the whole showing a more consistent pattern: There is more overlap in the semantic classes of verbs and a preference for inceptive readings for its top verbal collocates. The subjects typically refer to humans or institutions. With *quick to point out*, it also has a very dominant most frequent member whose inceptive reading might serve as an analogical model for the other verbs occurring in this construction.

Note that the only distributional difference discussed in all of the above that does not require careful semantic annotation is the contrast involving the *of*-phrases in the simple predicative pattern. Little is known about the distribution of these four pairs across the patterns in actual corpora, and one main goal of this paper is to explore distributional semantics as a means to close in on the specifics of the different items, ideally being able to link the results of the distributional analysis to ideas discussed in the theoretical literature. As a side effect, this might also shed some light on an unresolved issue regarding the relationship between the base and the

-ly forms: is it derivation or inflection? Bauer et al. (2013: 536) write “that the evidence is inconclusive”.

3 The target patterns in the ukWaC

The ukWaC is a 2 billion word corpus of English (Baroni et al. 2009). It is web-derived from only the .uk domain, therefore likely to be more representative of British English than any other variety. The version I used was part-of-speech-tagged and lemmatized with TreeTagger. Each adjective and the corresponding *-ly* form are treated as separate lemmata. All eight lemmata are high-frequency items, as seen in Table 1, and their distribution across the target patterns is shown in Table 2.

pair	base form	<i>-ly</i> form	total
<i>quick/quickly</i>	135705	187725	323430
<i>slow/slowly</i>	84391	59272	143663
<i>wise/wisely</i>	37950	6730	44680
<i>lucky/luckily</i>	51252	13911	65163

Table 1 Raw frequencies adjective vs. adverb overview

	quick	slow	wise	lucky
attrib	72607	33791	11551	12309
INF	8035	4733	5280	6034
pred	12307	15884	4563	20220
advSI	1433	2475	236	8950
advV	64149	22600	1848	513
Vadv	36178	15262	2405	64

Table 2 Distribution across target patterns

Within the adjective forms of *quick*, *slow*, and *wise*, the attributive usage is the most frequent usage. For *lucky*, the attributive usage comes second after the standard predicative usage. For *wise*, the *to*-INF pattern is actually slightly more frequent than the standard predicative usage. Closer inspection reveals that the *wise-to*-INF pattern is special in that there is a high proportion of instances of subject extraposition in the data, as illustrated in (19).

- (19) a. So, if you have one then it is wise to send it so we can add it to the site.
 b. But it is wise to be on your guard against these abuses.

Excluding all instances of *it* followed by any form of *be* from the *to*-INF pattern halves the instances for *wise*, reduces the instances of *quick* by 20% and leaves *slow* and *lucky* by and large unchanged, as seen in Table 3.

	quick	slow	wise	lucky
INF	8035	4733	5280	6034
INF (excluding <i>it</i>)	6308	4556	2650	6004

Table 3 The ADJ-*to*-INF pattern including and excluding *it* in subject position

While the exact reason for the preponderance of this pattern for *wise* and the effect of excluding this type on the distributional analysis must await further investigation, it is clear that the corresponding sentences are not paraphrasable by sentences with sentence-initial *wisely*. I therefore excluded this subpattern from further analysis.

For the adverb forms, it is noticeable that both *quickly* and *slowly* most often occur preverbally, followed by their postverbal usage. The sentence-initial usage is the least frequent. In contrast, *wisely* occurs more often postverbally than preverbally. *Luckily* is special in being the only item occurring most frequently in sentence-initial position, and rarely preverbally and even more rarely postverbally. This behavior of *luckily* is in line with the observation in the theoretical literature that it does not allow a low reading, which is syntactically associated with preverbal and postverbal position.

4 Distributional analysis

The distributional analysis compares the similarities of the four adjective/adverb pairs across three adjectival and three adverbial patterns. The three adjectival patterns are (a) the adjective in attributive position, (b) the adjective in predicative position followed by *to*-INF (excluding the subpattern in which the auxiliary is immediately preceded by *it*, following the discussion above), and (c) the adjective in predicative position not followed by *to*-INF. The three adverbial patterns are (a) the *-ly* form in

sentence-initial position, (b) the *-ly* form immediately preceding a main verb, and (c) the *-ly* form in postverbal position.

4.1 Preliminaries

To compare the different usages of each adjective, I used distributional semantics. The main idea behind this approach is to represent words exclusively via their distribution (for a comprehensive introduction and overview, see Sahlgren 2006). There are many different ways this can be done, and I proceeded as follows:

4.1.1 Initial steps

1. I first collected cooccurrence counts for each adjective, distinguishing between the three adjectival and the three adverbial patterns in the tagged ukWaC corpus.
2. The cooccurrence counts were collected for the top 10,000 content words (nouns, verbs, adjectives, and adverbs).
3. Only cooccurrences of at least 5 were used in the further calculations.

4.1.2 Further parameter setting and validation

Lapesa & Evert (2014: 542) point out that the three parameters score, transformation, and distance metric consistently play a crucial role in the performance of distributional semantic models, while the parameter of window size is influential but more task dependent. To find a useful setup for the task at hand, I first compared the performance of different settings against the human ratings on the adjective subset of the SimLex-999 dataset described in Hill et al. (2014). This dataset provides two human judgements of interest: a similarity score and an association score. For the similarity score, raters were instructed to rate synonyms and near-synonyms high, and to not confuse similarity with relatedness. For example, the pair *glasses/spectacles* was given as a reference for a pair with very similar meanings, whereas pairs like *car/tyre*, *car/crash*, and *car/motorway* were used to exemplify the difference between similarity and semantic relatedness. The ratings were given by setting a slider on an integer scale ranging from 0 to 6 with the low values indicating “less similar” and the high values indicating “more similar”. In the dataset itself, these values are linearly mapped from the [0,6] range to the [0,10] range. The association

score gives the strength of free association from one word in a pair to its partner. Values are taken from the University of South Florida Free Association Dataset (Nelson et al. 1998). In particular, they use the cue-to-target strength, arrived at by dividing the number of participants producing the target by the number of participants seeing the cue. In the SimLex dataset, this value is multiplied by the factor 10.

While I only used cosine similarity as a distance metric, I used a Python script to vary the scoring and transformation, exploring pointwise mutual information and log-likelihood, both with and without logarithm, and the window size, considering complete sentences, and ranges of two to four words to the left and to the right of the target word. The best performing settings for the similarity scores was a window of three words to the left and right of the target and pointwise mutual information without logarithm. The best performing setting for the association scores was a window of two words, and pointwise mutual information (pmi). Both are shown below, with the best correlation score for each of the two measures in boldface, and the best performing settings for the sentence window at the bottom.

window	score and trans	similarity		association	
		cor	p value	cor	p value
3	pmi w/out log	0.518	5.6e-09	0.219	0.021
2	pmi	0.452	6.5e-07	0.359	1.2e-04
sentence	pmi w/out log	0.331	3.9e-04	0.304	0.001

Table 4 Best performing settings overall and best performing setting with a sentence window

As Table 4 shows, both narrow window versions clearly outperform the sentence window on the similarity task, and a two word window version also outperforms the sentence window version on the association task. So far, only base forms of different adjectives were compared. For many adjectives, these forms share a preference for attributive position. In contrast, the adjective and the adverb of each pair typically never occur in the same position, and I further distinguish three distinct syntactic environments for each form. To further explore which setting to use for the comparison of these pairs across the syntactic environments, I calculated

the similarity score between the adjective and the adverb of each of the four pairs using the three settings. This reveals a stark difference between the window sizes, as shown in Table 5.

pair	sentence	three words	two words/pmi
<i>quick/quickly</i>	0.63	0.13	0.24
<i>slow/slowly</i>	0.63	0.16	0.29
<i>wise/wisely</i>	0.28	0.04	0.09
<i>lucky/luckily</i>	0.46	0.08	0.20

Table 5 Cosine similarities between adjective and adverb of each pair, using the best settings per window size

A crucial factor responsible for these differences is the apparent inability of the narrow word windows to meaningfully represent the similarity between the attributive usage and the usages of the *-ly* forms. For example, the similarities between attributive *quick* and all usages of *quickly* are only at 0.07 and 0.06 for the two word and three word windows in the above settings, and not higher than 0.04 for the three adverbial positions considered here individually. In contrast, the similarity between attributive *quick* and all *-ly* forms is 0.52 when using the sentential window in its best performing settings.

One factor behind this becomes clear when considering textbook equivalences of adjective and adverb usages illustrated in (20).

- (20) a. The quick runner ...
 b. The man runs quickly.

While the head noun *runner* in (20a) is an important and straightforward cue for distributional systems that compare different adjectives, it is not as straightforwardly helpful in establishing a similarity across the two sentences, because the system is blind to the relationship between the verbal base *run* and its nominalization *runner*. The larger contexts used by the sentence window seems to be able to circumvent this problem.

Since the sentence-based window is better suited to comparisons across constructions and forms, it is used when comparing pairs across usages. When forms are compared within a pattern, I will use the setting perform-

ing best for the similarity task, that is, a 3-word window with pmi without log. Note that this method does not reduce the effects of absolute frequency in the corpus completely. Given the considerable differences in absolute frequencies between the six patterns and the four pairs, this factor might well play a role. However, of all 24 pattern/form combinations, only postverbal *luckily* is a clear low frequency outlier.

4.2 Results

The resulting similarities are shown for each of the four adjectives in Tables 6–9. A cosine similarity of 1 indicates perfect similarity (the vectors point in the same direction). The closer the value gets to 1, the more similar two vectors are. A cosine of 0, corresponding to a 90 degree angle, indicates unrelated scores. Negative cosine values are not possible in the two setups selected in §4.1.2, as both use pmi without log on count data which cannot result in negative values. NA in the last column of Table 9 results from the rarity of postverbal *luckily*; the corresponding similarities could not be meaningfully calculated.

pattern	INF	pred	advSI	advV	Vadv
attrib	0.20	0.33	0.09	0.49	0.41
INF		0.16	0.10	0.33	0.26
pred			0.09	0.33	0.31
advSI				0.16	0.12
advV					0.68

Table 6 Cosine similarities between *quick*-usages.

pattern	INF	pred	advSI	advV	Vadv
attrib	0.20	0.57	0.14	0.49	0.48
INF		0.28	0.08	0.19	0.19
pred			0.12	0.34	0.36
advSI				0.45	0.29
advV					0.73

Table 7 Cosine similarities between *slow*-usages.

Looking at the patterns across the four items, the tables show the fol-

pattern	INF	pred	advSI	advV	Vadv
attrib	0.17	0.51	0.02	0.17	0.13
INF		0.20	0.09	0.17	0.13
pred			0.01	0.16	0.13
advSI				0.13	0.03
advV					0.17

Table 8 Cosine similarities between *wise*-usages.

pattern	INF	pred	advSI	advV	Vadv
attrib	0.21	0.33	0.19	0.06	NA
INF		0.41	0.34	0.22	NA
pred			0.40	0.14	NA
advSI				0.18	NA
advV					NA

Table 9 Cosine similarities between *lucky*-usages.

lowing:

1. The attributive usage across all four adjectives shows little similarity to *to*-INF, with all similarities in the narrow band between 0.17 and 0.21. The attributive usage is more similar to the standard predicative usages, with a clear difference between *slow* and *wise* (0.57/0.51), on the one hand, and *quick* and *lucky*, on the other hand (0.33/0.33). The highest similarity to the sentence-initial adverbial pattern is only 0.19, for *lucky*. The similarities to the pre- and postverbal adverbs show a clear split between relatively high values for *quick* and *slow*, on the one hand, and the other two adverbs, which show little or no similarity, on the other hand.
2. The *to*-INF pattern varies a lot in its similarity to the other usages across the four adjectives. With the exception of *lucky*, it is less similar to the predicative pattern than the attributive is to the predicative pattern. *Lucky* also stands out when comparing the similarity between the *to*-INF pattern and the sentence-initial adverb: the cosine value is 0.34, as opposed to no or hardly any similarity for the other three items. The similarity between *to*-INF and the preverbal adverbs is at the same level for *slow/wise*, and *lucky*, with a higher

value for *quick* (0.33). *Quick* has also the highest similarity between INF and the postverbal adverb (0.26), with only 0.19 and 0.13 for *slow* and *wise*, respectively.

3. The similarity between the predicative usage and the sentence-initial adverbs is relatively high for *lucky* (0.41), low for *quick/slow*, and non-existing for *wise*. It is moderately high for preverbal and postverbal *quick* and *slow*, and low for *wise* and preverbal *lucky*.
4. The sentence-initial pattern is always more similar to the preverbal usages than the postverbal ones, with moderate similarity to the preverbal and postverbal adverb for *slow*, and lower similarities with the preverbal adverb for *quick/wise/lucky*, and little or not at all similar for postverbal *quickly* and *wisely*.
5. The preverbal and postverbal adverbs are highly similar for *quickly/slowly* (0.68/0.73), showing little similarity (0.17) for *wise*. As mentioned above, for *lucky* this contrast does not apply.

5 Discussion

This section cannot meaningfully discuss all the nuances of the distributional data across the four adjective/adverb pairs. Instead, I will focus on contrasts that are of special interest in view of the discussions and classifications in the previous literature. Therefore, I will discuss the following points more closely:

1. Event and non-event predicates
2. The *to*-INF pattern and the adverbial usages
3. The preverbal and postverbal adverb positions
4. The relationship between forms with *-ly* and without *-ly*

5.1 Event and non-event predicates

In the literature, *quick* and *slow* are held to be event predicates, as opposed to *wise* and *lucky*. *Wise* is claimed to make covert reference to events. Of the adverbial positions, both the preverbal and the postverbal position allow low readings, which directly relate to events. The preverbal position also allows high readings, which, for *quick* and *slow* at least, are both linked to the event encoded by the verbal predicate (see §5.3 for more discussion). Since *quick* and *slow* are event predicates, they can simply be used in the same way there as in attributive and predicative position. All

four usages should be similar to each other. In contrast, for *wise*, the low adverbial reading should be quite different from its attributive and predicative usage, because now it must be brought in a direct relation with an event. Table 10 shows the similarities for the relevant patterns.

adjective	attrib		predicative	
	advV	Vadv	advV	Vadv
<i>quick</i>	0.49	0.41	0.33	0.31
<i>slow</i>	0.49	0.48	0.34	0.36
<i>wise</i>	0.17	0.13	0.16	0.13
<i>lucky</i>	0.06	NA	0.14	NA

Table 10 Cosine similarities between the attributive and the predicative patterns and the preverbal and postverbal patterns for all four pairs

The observed similarities are by and large in line with the expectations: both *quick* and *slow* have relatively high similarity values for both preverbal and postverbal adverbial usages, while for *wise*, the values are markedly lower. While it is unclear how many of the preverbal usages correspond to the high reading, the similarity value here is still not very high for *wise*, showing that it is more distinct across its usages overall.

Note that the values for *lucky* are also low. I do not have an explanation for this. It might have to do with the preverbal usage of *luckily* being relatively rare and perhaps restricted to more idiosyncratic combinations.

Another quantitative effect of *wise* behaving more like an event predicate in the two adverbial usages can be seen when looking at the similarities between *wise* and *quick/slow* across the respective usages. Table 11 shows the similarities between the relevant four different usages of *wise* to the corresponding usages of *quick* and *slow* (because I am now comparing the same usage, the reported similarities are from the best performing similarity setting, that is, using a 3-word window and pmi without log).

While there is no or almost no similarity across the adjective usages, similarity slightly increases in the preverbal adverbial usage. There is a marked increase in similarity for the postverbal usage, that is, the position restricted to low readings.

across-item similarity: <i>wise</i>				
adjective	attributive	predicative	preverbal	postverbal
<i>quick</i>	0.00	0.01	0.02	0.05
<i>slow</i>	0.00	0.00	0.01	0.12

Table 11 Cosine similarities between four *wise*-patterns and the corresponding *quick/slow* patterns.

5.2 The *to*-INF pattern and the adverbial usages

It is unclear what kind of similarities to expect for the *to*-INF pattern in comparison to the attributive and predicative usage. For *quick*, I argue in Schäfer (2020a) that its occurrence in the *to*-INF pattern is maximally different from its usage in the attributive position. Among other things, it is restricted to intentionally acting subjects, typically, humans or institutions. This would lead one to expect a low similarity between these two usages, and a similar point could be made for *slow*. More interesting are the expectations for the similarities to the adverbial readings, especially those described as roughly synonymous. Since the *to*-INF construction is aligned with the higher readings, I expect it to be more similar to the sentence-initial and preverbal occurrences of *quick*, *slow*, and *wise* than to the postverbal one. For *luckily*, there is only one reading, but the restriction to subject-oriented interpretations might make it less similar than the correspondences to the high readings for the other adverbs. I have no clear idea what influence the presupposition/assertion reversal for *wise* should have; see below for more discussion.

pattern	attrib	pred	advSI	advV	Vadv
<i>quick-to</i> -INF	0.20	0.16	0.10	0.33	0.26
<i>slow-to</i> -INF	0.20	0.28	0.08	0.19	0.19
<i>wise-to</i> -INF	0.17	0.20	0.09	0.17	0.13
<i>lucky-to</i> -INF	0.21	0.41	0.34	0.22	NA

Table 12 Cosine similarities between the *to*-INF pattern and all other patterns for each pair

Among the similarities of the *to*-INF pattern to the other usages, as seen in Table 12, the relation to the attributive usages stands out as being the

most consistent in magnitude. For all four adjectives, it is also clearly lower than the corresponding attrib-pred values (0.33, 0.57, 0.51, 0.33). *Lucky-to*-INF/*luckily* differ from the three corresponding pairings due to their high value for the sentence-initial usage. The values for the similarity to the preverbal usage are relatively consistent for all four adjectives, with *quick* an outlier in being markedly more similar. The three values for the postverbal usage are all different by at least 0.06, and there is a clear drop in similarity for *quick* and *wise*, but not for *slow*.

All this only partially matches the expectations. As expected, *quick-to* and *slow-to* are less similar to their corresponding attributive usages than the attributive usage is to the predicative usage. However, this also holds for *lucky* and *wise*. As for the adverbial usages, the high value for sentence-initial *luckily* is expected, the consistently low values for the other three adverbs are unexpected. The most likely explanation for these low values is that the sentence-initial position is only a very marginal option for these adverbs, used when regular options are exhausted. This is in line with the observation that it is, for all three adverbs, the least frequent usage, less frequent than their preverbal and postverbal usages by at least a factor of 6. In contrast, for *luckily*, the sentence-initial pattern is the dominant pattern, with the preverbal and postverbal usages less frequent by more than a factor of 10 and of 100, respectively. That the similarity for *luckily* is also the highest overall is unexpected giving that the usages are restricted to subject-oriented interpretations (more on this in §5.2.1). The drop in similarity between preverbal and postverbal adverbs observed for *quick* and *wise* is expected. Why does it not obtain for *slow*? It could be related to the observation in Schäfer (2020b) that *quick-to*-INF is internally more coherent, but further research is needed here.

5.2.1 The subject-oriented reading of *luckily*

The narrowing of the interpretation of *lucky* to an evaluation relative to the subject seems not to be something that is picked up by the distributional analysis. In fact, the sentence-initial similarity value is the highest across all four adjectives. Why might this be? I think that the analysis of *luckily* as a speaker-oriented evaluative is correct, that is, it is the speaker's evaluation of a fact. For who this is lucky can be made explicit by a *for*-phrase, cf. (21).

- (21) Luckily for her, she had a clever lawyer at her trial, and was never punished for the murder. [FPU 1866]

Without an explicit *for*-phrase, the tendency seems to be to assume that it is lucky for the speaker, as in (22), where the context does indeed confirm that it was an instance of luck for the speaker, and what it meant for the referent of *she* is irrelevant.

- (22) Luckily she left the school! [KA1 2120]

My impression is, though, that, when combined with sentences with human subjects, the speaker's evaluation often plausibly coincides with the assessment the speaker would give when taking the viewpoint of the subject, as in a typical example for *luckily* (from Ernst 2009) in (23).

- (23) a. Luckily, Aaron did not fall off his bicycle. (= Ernst's (1c))
 b. It is lucky that Aaron did not fall off his bicycle. (= Ernst's (2b))

While Ernst gives (23b) as a paraphrase, the corresponding *to*-INF sentence will in most instances also be consistent with the situation described in (23a), as shown in (24).

- (24) Aaron was lucky to not fall off his bicycle.

Note that these subject-oriented readings are still different from the standard examples of subject-oriented adverbials in not describing a mental attitude, nor are they agent-oriented in the sense that the agent has control over the action (the notion of agent orientation in Ernst 2002: 55). That is, the readings are best described as still corresponding to speaker-oriented adverbials, with the target of the luck made explicit, similar to the adverbial usage with a *for*-phrase. If this is correct, then the difference is indeed a minor one, and perhaps expected to not impact much on the distributional similarity measure.

5.2.2 The assertion/presupposition reversal for wisely

The correspondence between *wisely* and *wise-to* INF is on average the lowest. One reason could be that the distributional analysis is able to pick up on the assertion/presupposition reversal. However, this is impossible

to ascertain when one looks at other work on entailments in distributional semantics.

Distributional semantics is in principle able to deal with entailments, and there are numerous studies dealing with lexical entailment (cf. Turney & Mohammad 2014 for a comparison of different approaches). Studies on entailment above the word level also exist, but the domain for which entailment in a strict formal semantic sense is explored is typically still very small (cf. Baroni et al. 2012, who report two experiments on adj-noun to noun and quantifier phrase to quantifier phrase entailments). At the sentence level, the distinction between entailment and presupposition seems to be largely irrelevant for the NLP community, where this task falls into the domain of recognizing textual entailment: “In RTE [recognizing textual entailment], the gold standard for entailment is established by common sense, rather than formal logic” (Turney & Mohammad 2014: 2). No differentiation is usually made between the role of linguistic and world knowledge in this. In general, the inferences are not seen as absolute, and a well-known approach states “We say that T entails H if, typically, a human reading T would infer that H is most probably true” (Dagan et al. 2009: iv). That is, there is also no specific distinction between entailment and implicature. However, on this level, all four adjectives pattern together, as opposed to adjectives like *eager* in this construction.

- (25) It was a huge budget and the top agencies were eager to get their hands on it. [ADK 667]

This lack of interest in the distinction between entailments, implicatures, and presuppositions seems to be driven mostly by the assumption that this level of detail is irrelevant for applied tasks. I don't know of any study that uses training data distinguishing between presuppositions and implicatures. Either way, I have no bottom line to compare the possible effect of *wise/wisely* to, since, in my data, *wisely* is the only adverb that has the standard subject-oriented adverbial reading. Its presuppositional behavior is thus not its only difference to the three other adverbs.

5.3 The preverbal and postverbal adverb positions

The literature assumes, where available, pairs of high and low readings for the adverbs. Of the positions tested, the sentence-initial position is re-

served for high readings, and the postverbal position for low readings. The position immediately preverbal is ambiguous between high and low readings when no auxiliaries are present, whose presence was not controlled for in this study. For *quickly* and *slowly*, this means that postverbally only true rate readings are predicted to occur, with the immediately preverbal position in addition allowing inceptive and holistic readings. For *wisely*, only the manner reading is expected to occur postverbally, while the high reading is restricted to preverbal position. Leaving aside the issue of how many readings in the preverbal position are actual high readings, the two variants are more closely related for *quick* and *slow*. This not only holds on the theoretical level, but also conceptually: the different readings often stand in implicative relationships. This can already be seen when looking at (10) from §2, repeated here for convenience:

- (26) a. Lynn quickly raced down the hallway.
 b. Lynn raced down the hallway quickly.

For Ernst, (26a) “can be interpreted as saying that Lynn’s beginning the action of racing-down-the-hallway occurred quickly after some other event (perhaps a command to go fetch something), while (2.149b)[(26b)] is a description of the speed of her movement” Ernst (2002: 85). I suspect that in real world situations, the inceptive interpretation of (26a) typically is taken to imply the true rate reading of (26b). And given that the hallway is of a limited length, the true rate reading for (26b) implies the holistic interpretation that the whole action only took a short amount of time. In this particular example, since *to race* already is something connected to high speed, the true rate reading might be even more likely to imply the holistic interpretation. This requires more investigation, as it is also clear that these implicative relationships do not obtain for all verb types (for example, *quickly* in combination with stative verbs cannot receive a true rate reading and typically receives an inceptive reading). In contrast, the high and low reading of subject-oriented adverbs like *wisely* are conceptually clearly distinct, and, while not very common, opposites can be used with the respective readings, leading to much-discussed examples like Parson’s (1972) *John painstakingly wrote illegibly*. Thus, we can expect higher similarity values for *quickly* and *slowly* between the preverbal and postverbal

usages, and lower ones for *wisely*. This is borne out by the data; in fact, the preverbal and postverbal usages of *quickly* and *slowly* lead to the highest similarity values observed in this study, 0.68 and 0.73, respectively. In contrast, the corresponding value for *wisely* is a mere 0.17.

5.4 The relationship between forms with *-ly* and without *-ly*

The relationship between the adjectival base form and the *-ly* form is often discussed in English linguistics. The issue is whether it should best be seen as a derivational suffix or an inflectional suffix: see Payne et al. (2010) and Giegerich (2012) as representatives of the two positions.

Criteria that can be approached by distributional semantics should concern meaning. Plag (2003: 195–196) argues that two meaning-related aspects have a bearing on this issue: *-ly* does not encode lexical meaning, which would be expected from derivation, and *-ly* is always semantically transparent (the latter with only a few exceptions), which would be expected from inflection.

Bonami & Paperno (2018) discuss a list of five criteria from Stump (1998). These also includes Plag's two points (instead of semantic transparency, Stump speaks of semantic regularity: "inflection is semantically more regular than derivation"). They point out that these criteria are "formulated in terms of high-level morphological notions that are not easy to operationalize" and instead provide an operationalization of the semantic regularity criterion in terms of stability of contrasts, as given in (27) (= Bonami & Paperno's (2)):

- (27) **Stability of contrast:** The morphosyntactic and semantic contrasts between pairs of words related by the same *inflectional* relation are more similar to one another than the contrasts between pairs of words related by the same *derivational* relation.

Bonami & Paperno explored this criterion by looking at sets of triplets of <pivot, inflectionally related form, derivationally related form> in French. For example, one such triplet was <verb INF, verb PST.IPFV.3SG, SG *-eur* noun>, e.g. <*baigner* 'to bathe', *baignait* 'bathed', *baigneur* 'bather'>. They found that overall the contrasts between inflectionally related forms were more stable (measured in terms of dispersion around the average vector offset for a specific paradigmatic system).

While their approach was word-form based, the comparison across patterns here used lemmatized forms, and I only looked at four word pairs. However, even these four pairs are interesting because they suggested that a combination of word form and syntactic pattern can yield more insight into this issue. Recall that the overall similarity between the two forms across all usages were 0.63, 0.63, 0.28, and 0.46, respectively, leading to an overall standard deviation of 0.17. Taking the standard deviation as a measure of dispersion, we can now compare the standard deviations across all usages, as seen in Table 13.

pattern	INF	pred	advSI	advV	Vadv
attrib	0.02	0.12	0.07	0.22	0.23
INF		0.11	0.13	0.07	0.11
pred			0.17	0.10	0.16
advSI				0.15	0.13
advV					0.36

Table 13 Standard deviations of the four similarity values for the four pairs in each pairing

As the table shows, the dispersions for the similarities within the adjective usages are all clearly below 0.17, with 0.12 the highest value. Within the adverbial usages, the standard deviation for the similarities between preverbal and postverbal usage is 0.36, the highest standard deviation in the table. The two other values, while higher than any of the adjectival values, fall below the threshold of 0.17. Looking at the values for similarities involving one adjectival and one adverbial usage, we see a wide range, with only two values higher than 0.17. Only four out of nine values fall in the range of the adjective-only values: the standard deviations for the similarities between attributive adjective and sentence initial adverb, *to*-INF and preverbal adverb, *to*-INF and postverbal adverb, and simple predicative usage and preverbal adverb. Following the general logic of Bonami & Paperno (2018), these values support the assumption that *-ly* is derivational: low dispersion around different usages not involving derivation, higher values between derived forms, and on average higher values for pairs of forms related by derivation.

6 Summary and outlook

This paper presented a pilot study using distributional semantics to compare four adjective/adverb pairs across six different usages. The four pairs were meant to represent a variety of different lexical types that are discussed in the literature to be reflected in their linguistic behavior. The general idea was to explore whether qualitative observations made about these items and their corresponding classes show up in their distributional characteristics. After presenting the distributional analysis of *quick*, *slow*, *wise*, and *lucky* across three adjectival and three adverb patterns, four areas were discussed in more detail:

1. The assumption that *quick* and *slow* are event predicates is in line with the high similarity across their standard adjectival and standard adverbial usages, contrasting with *wise*, which shows consistently lower similarities.
2. The comparison between the *to*-INF pattern and the adverbial patterns only partially reflected expectations. In general, *lucky* behaved differently from the three other adjectives, which might in part be due to its already very different distribution of absolute frequencies. Also, all three other adjectives were expected to be more similar to the preverbal usages, but this was only the case for *quick* and *wise*, and not for *slow*. Further, it remained unclear whether anything in the distributional data was sensitive to the subject-oriented interpretation of *lucky* in the INF construction and the presupposition/assertion reversal observed for the *wise-to*-INF construction.
3. Of the three adverbs that occurred frequently enough in both pre- and postverbal position, *quickly* and *slowly* were relatively similar, in contrast to *wisely*. This is in line with the observation that the high and low readings for *wisely* are clearly conceptually distinct. For *quickly* and *slowly*, the respective readings in many cases imply at least one of the other readings.
4. The data presented is in line with the assumption that *-ly* is a derivational affix if one assumes that derivation is linked with less stability of contrasts, here operationalized by higher standard deviations across the similarity values for a given comparison.

Overall, this paper has shown that the combination of distributional se-

mantics of different modifier usages with qualitative analysis is a promising step forward in the analysis of the semantics of adjectives and adverbs and their interrelation. Many findings that remained unresolvable here might be resolvable when more pairs are taken into account. Consideration of a larger number of pairs, and also exploration of further distributional setups, would also allow one to clarify the possible influence of the raw frequencies of the forms across the six patterns on the results reported here.

Acknowledgments Many thanks to the original anonymous reviewers of my abstract for CSSP and to the audience in Paris for their precise and helpful comments. Heartfelt thanks also to one anonymous reviewer and Chris Piñón for their careful reading and comments on the final version of this paper. Work on this paper was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – SFB 833 – Project ID 75650358.

References

- Barker, Chris. 2002. The dynamics of vagueness. *Linguistics and Philosophy* 25(1). 1–36. doi:10.1023/A:1014346114955.
- Baroni, Marco, Raffaella Bernardi, Ngoc-Quynh Do & Chung-chieh Shan. 2012. Entailment above the word level in distributional semantics. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*. 23–32. Avignon, France: Association for Computational Linguistics. <https://www.aclweb.org/anthology/E12-1004>.
- Baroni, Marco, Silvia Bernardini, Adriano Ferraresi & Eros Zanchetta. 2009. The wacky wide web: a collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation* 43(3). 209–226. doi:10.1007/s10579-009-9081-4.
- Bauer, Laurie, Rochelle Lieber & Ingo Plag. 2013. *The Oxford reference guide to English morphology*. Oxford: Oxford University Press.
- Bonami, Olivier & Denis Paperno. 2018. Inflection vs. derivation in a distributional vector space. *Lingue e Linguaggio* 17(2). 173–195.
- Bücking, Sebastian & Claudia Maienborn. 2019. Coercion by modification. The adaptive capacities of event-sensitive adnominal modifiers. *Semantics and Pragmatics* 12(9). doi:10.3765/sp.12.9.
- Cinque, Guglielmo. 1999. *Adverbs and functional heads: A cross-linguistic perspective*. Oxford: Oxford University Press.

- Cruse, Alan. 2004. *Meaning in language: An introduction to semantics and pragmatics*. Oxford: Oxford University Press.
- Dagan, Ido, Bill Dolan, Bernardo Magnini & Dan Roth. 2009. Recognizing textual entailment: Rational, evaluation and approaches. *Natural Language Engineering* 15(4). i–xvii. doi:10.1017/S1351324909990209.
- Dik, Simon C. 1975. The semantic representation of manner adverbials. In A. Kraak (ed.), *Linguistics in the Netherlands 1972–1973*, 96–121. Assen: Van Gorcum.
- Ernst, Thomas. 2002. *The syntax of adjuncts*. Cambridge: Cambridge University Press.
- Ernst, Thomas. 2009. Speaker-oriented adverbs. *Natural Language and Linguistic Theory* 27. 497–544.
- Geuder, Wilhelm. 2002. *Oriented adverbs: Issues in the lexical semantics of event adverbs*. Tübingen: Universität Tübingen dissertation.
- Giegerich, Heinz J. 2012. The morphology of *-ly* and the categorial status of ‘adverbs’ in English. *English Language and Linguistics* 16(3). 341–359. doi: 10.1017/S1360674312000147.
- Hill, Felix, Roi Reichart & Anna Korhonen. 2014. Simlex-999: Evaluating semantic models with (genuine) similarity estimation.
- Karttunen, Lauri. 2013. You will be lucky to break even. In Tracy Holloway King & Valeria de Paivaw (eds.), *From quirky case to representing space: Papers in Honor of Annie Zaenen*, 167–180. Stanford, CA: CSLI Publications.
- Lapesa, Gabriella & Stefan Evert. 2014. A large scale evaluation of distributional semantic models: Parameters, interactions and model selection. *Transactions of the Association for Computational Linguistics* 2. 531–545.
- Larson, Richard K. 1998. Events and modification in nominals. In D. Strolovitch & A. Lawson (eds.), *Proceedings from Semantics and Linguistic Theory* 8. Ithaca, NY: Cornell University.
- Maienborn, Claudia & Martin Schäfer. 2011. Adverbials and adverbs. In *Semantics: An international handbook of natural language meaning*, vol. 2, 1390–1420. Berlin: Mouton de Gruyter.
- Nelson, D. L., C. L. McEvoy & T. A. Schreiber. 1998. The University of South Florida word association, rhyme, and word fragment norms. <http://www.usf.edu/FreeAssociation/>.
- Oshima, David Y. 2009. Between being wise and acting wise: A hidden conditional in some constructions with propensity adjectives. *Journal of Linguistics* 45(2). 363–393. doi:10.1017/S0022226709005684.
- Parsons, Terence. 1972. Some problems concerning the logic of grammatical

- modifiers. In Donald Davidson & Gilbert Harman (eds.), *Semantics of natural language* (Synthese Library). 127–141. Dordrecht: D. Reidel.
- Payne, John, Rodney Huddleston & Geoffrey K. Pullum. 2010. The distribution and category status of adjectives and adverbs. *Word Structure* 3(1). 31–81.
- Piñón, Christopher. 2008. From properties to manners: A historical line of thought about manner adverbs. *Studies van de Belgische Kring voor Linguïstiek – Travaux du Cercle Belge de Linguistique – Papers of the Linguistic Society of Belgium* 3.
- Plag, Ingo. 2003. *Word-formation in English*. Cambridge: Cambridge University Press.
- Pustejovsky, James. 1995. *The generative lexicon*. The MIT Press.
- Sahlgren, Magnus. 2006. *The word-space model: Using distributional analysis to represent syntagmatic and paradigmatic relations between words in high-dimensional vector spaces*. Stockholm: Stockholm University dissertation.
- Schäfer, Martin. 2013. *Positions and interpretations: German adverbial adjectives at the syntax-semantics interface*. Berlin: De Gruyter Mouton.
- Schäfer, Martin. 2020a. From *quick* to *quick-to*-INFINITIVAL: On what is lexeme specific across paradigmatic and syntagmatic distributions. *English Language and Linguistics* 1–31. doi:10.1017/S1360674320000167.
- Schäfer, Martin. 2020b. Modifier semantics and constructional idiosyncrasies in the English ADJ-*to*-INF construction. Submitted to an edited volume on Modifiers and Complex predicates.
- Stump, Gregory T. 1998. Inflection. In A. Spencer & A. Zwicky (eds.), *The handbook of morphology*, 13–43. London: Blackwell.
- Turney, P. D. & S. M. Mohammad. 2014. Experiments with three approaches to recognizing lexical entailment. *Natural Language Engineering* 21(3). 437–476. doi:10.1017/s1351324913000387.
- Wilkinson, Robert W. 1970. Factive complements and action complements. In *Proceedings of the Chicago Linguistic Society. Papers from the Sixth Regional Meeting*. 425–44. Chicago: Chicago Linguistic Society.