

A critical look at partial acceptability in English and Polish

Sylwiusz Żychliński
Adam Mickiewicz University, Poznań
sylwiusz@amu.edu.pl

Abstract

This paper is primarily concerned with the role of partial acceptability (and its ‘flavors’) in English linguistic examples taken from the field of generative grammar. Partially acceptable judgments as a subset of acceptability judgments form a heterogeneous group that plays a significant role in linguistic research and yet the extent to which partial acceptability is employed in the linguistic literature has not been studied extensively. The first part of the paper discusses a number of relevant issues related to the use of partial acceptability, such as problems with the conceptualization of gradience or excessive creativity with regard to levels of acceptability present in generative papers. The second part of the paper reports on a meta-linguistic study of partial acceptability in a pre-selected corpus of generative papers on English, with the focus on the types of partial acceptability and their function. Lastly, the English results are compared with the results of a previously conducted similar study of partial acceptability in Polish.

Keywords: linguistic examples; informal acceptability judgments; partial acceptability; gradience; intuitions.

1. Introduction

Acceptability judgments (henceforth AJs), and various methods of obtaining linguistic data in particular, have been one of the hot topics in linguistics for more than two decades now. The main problem with AJs is stated in Cowart (1997: 1), who says that “[s]o far as research practice is concerned, one of the most striking contrasts between generative linguistics and its immediate predecessors lies in the fundamental importance assigned to speakers’ introspective judgments of sentence acceptability.” Schütze (1996: 1) adds that “[generative grammar] involves constructing theories of intuition rather than of

language use, and (...) it is highly subjective and biased by the views of the linguist.” Ever since these two influential publications, the informal nature of AJs has been a recurring theme of numerous linguistic papers. Although it is easy to find both proponents (e.g. Phillips 2009; Dąbrowska 2010; Culicover & Jackendoff 2010; Sprouse & Almeida 2012, 2017; Sprouse et al. 2013) and opponents (among others, Wasow & Arnold 2005; Gibson & Fedorenko 2010) of informal AJs, many linguists will agree that the use of empirical methods in data collection is becoming more widely observed. Nothing seems to suggest, however, that linguists will (or even should) abandon the informal AJs as the most efficient source of linguistic information.¹ Thus, a complete understanding of the role of informal AJs is still vital. The main aim of this paper is to examine the way informal AJs have been employed in the generative research on English. Then, the English data will be compared with the results of a similar study previously conducted for Polish. However, instead of looking at the whole spectrum of AJs, the scope of the analysis has been restricted to partially acceptable judgments, which are perceived as a vague category and their application has not been widely discussed in the literature.

It should be emphasized that this article remains neutral to the question whether informal linguistic intuitions are the optimal source of linguistic data. Instead, the emphasis is placed on the fact that informal AJs are still the most common type of linguistic data and as such, they merit serious consideration. Ideally, the inconsistencies and weaknesses in the use of informal AJs brought to the surface in this paper will stimulate further discussions and, ultimately, the generative field will work out a set of uniform guidelines that will ensure the improved quality of informal AJs.

The rest of the article has the following organization. Section 2 provides an overview of a number of relevant issues related to partial acceptability. In Section 2.1 the discussion focuses on inconsistencies related to how the gradient structure of informal AJs is not entirely compatible with the conventional, odd-numbered rating scales used in semi-formal and formal experiments. Section 2.2 recognizes the spurious creativity in creating multiple layers of acceptability that are mostly counterproductive and create unnecessary meta-linguistic clutter. Attempts to make acceptability judgments less informal are

¹ What is more, Schütze (2020) shows that judgments obtained in experiments are not necessarily more reliable than informal judgments simply by virtue of being experimental data. He discusses a range of additional experimenting procedures, which are not the standard practice in typical empirical experiments, that would have to be followed in order to ensure that the experimental data can be properly interpreted.

discussed in Section 2.3, along with examples calling for the development of more specific guidelines in this area. Section 2.4 raises the question of the unclear criteria which determine whether a judgment should be annotated as partially acceptable. Section 3 provides a detailed description of the study of partial acceptability in English.² For the purpose of the study a corpus of texts from English linguistic journals has been selected and analyzed with the aim of assessing the extent to which partial acceptability is employed in linguistic articles. The types and functions of partial acceptability are described in Sections 3.1 and 3.2, respectively. In Section 3.3, the English results are compared with the results of a similar study from Polish based on Żychliński (2020). Section 4 offers conclusions.

2. Selected issues related to the use of partial acceptability in linguistic papers

Meta-linguistic studies of the application of partial acceptability are not common, which may be the reason why there are hardly any discussions of a host of issues related to its implementation in linguistic papers. A selection of such issues has been subjected to a closer inspection in the following subsections.

2.1. Conceptualization of gradience in AJs

One of the underspecified aspects of AJs is the problem of the conceptualization of gradience. In the most clear-cut scenario there are only three levels of acceptability, illustrated in (1) with examples from Postal & Pullum (1982) cited in Sato (2012).

- (1a) I don't want it to become standard practice in this monastery to flagellate oneself in public.
- (1b) ?I don't want to flagellate oneself in public to become standard practice in this monastery.

² A clarification is needed here. Partial acceptability in English means that the partially acceptable judgments were made for English examples. Later in this article, when these results are compared with the data from Polish, it should be stressed that the Polish papers in question were also written in English, but dealt with Polish AJs.

(1c) *I don't wanna flagellate oneself in public to become standard practice in this monastery.

(Postal & Pullum 1982: 124)

Most people will agree that sharp contrasts between (1a–b) and (1b–c) fully justify each of the three levels of acceptability. However, the range of acceptability judgments covering the area between (1a) and (1c) can be far more complex. And thus, only by looking at the notation used to mark partially acceptable examples (i.e. ?, ??, *?), a picture with at least five levels of acceptability emerges, as illustrated in Figure 1.

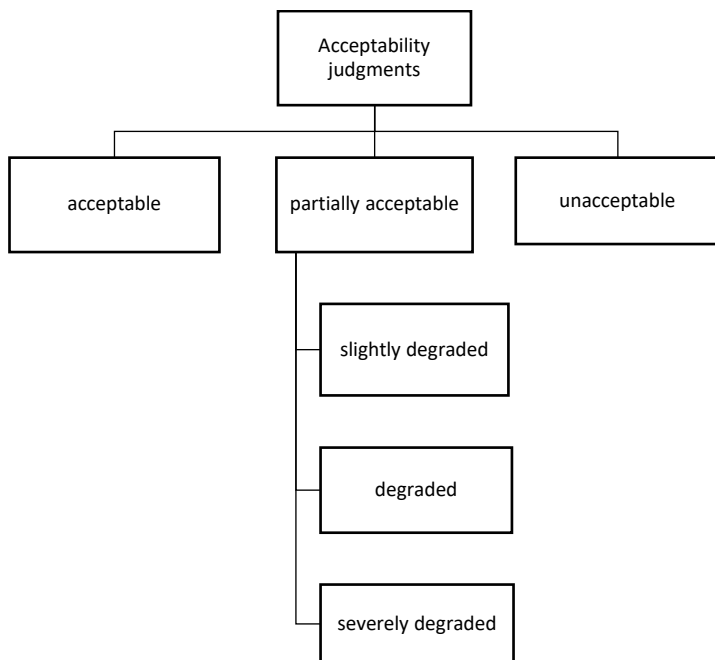


Figure 1. The structure of acceptability judgments.

Following conventional notation, examples which receive no additional notation represent judgments which are considered uniformly well-formed or fully

acceptable, as in (2). Slightly degraded judgments are marked with a question mark, as in (3), degraded judgments are distinguished by a double question mark, as in (4), a combination of a question mark and an asterisk indicates a severely degraded judgment, as in (5), and unacceptable judgments are prefixed with an asterisk, as in (6).

- (2) John and Bill sang and danced, respectively.
(Kubota & Levine 2016: 912)
- (3) ?John isn't inviting anyone without saying who he isn't inviting.
(Yoshida 2010: 350)
- (4) ??Each occasional beer is good for you.
(Gehrke & McNally 2015: 843)
- (5) ?*To what extent is that the moon is made of cheese a theory worth considering.
(Moulton 2013: 250)
- (6) *I saw two man without identifying themselves.
(Sobin 2014: 405)

This fine-grained structure is problematical for a few reasons. First and foremost, upon looking at a large enough corpus of linguistic examples, the assignment of the acceptability level comes across as a largely arbitrary process. It is surely somewhat motivated when two or more examples are grouped to show how their acceptability progressively degrades (see (15) in Section 2.2), but very often examples are either presented in isolation or alongside other examples that are completely different, which eliminates any reasonable basis for a meaningful comparison.

The second problem is that each of the five levels of acceptability is conceptualized differently by different people. This means that what constitutes slightly degraded level for one author may be completely different from another slightly degraded (based on the notation) example in another paper, as shown in (7–8). To make matters more complicated, in the case of example (8), Bruening (2010) adds that it was originally marked with a double question mark, but for him it is “close to perfect” (2010: 49).

- (7) ?Mary introduced me to herself.
(Sheehan 2013: 374)
- (8) ?To himself₁ is said to have been unexpectedly described the only
guy₁ who thought he was handsome.
(Bruening 2010: 50, after Postal 2004: 25)

These problems aside, it should be noted that in the standard notation used for informal AJs there is no middle ground, or neutral option, akin to ‘neither great nor completely bad’ or ‘so-so’. If at all, ? judgments may sometimes be described as ‘not perfect, but still passable’, but Figure 1 shows clearly that they are not in the center of the acceptability scale, with only acceptable judgments to their left but degraded, severely degraded and unacceptable judgments to their right. Why is all this relevant? The study of partial acceptability in linguistic articles has shown that there is no uniformity in the conceptualization of gradience when informal judgments are replaced with results obtained in empirical tests. As is well-known, the Likert-scale, which is probably the most commonly used rating scale in linguistic questionnaires, is an odd-numbered scale and regardless of whether a 5-point or 7-point rating is used, it implies that the mid-value is neutral-sounding, thus ‘not great but not terrible’. What does it mean for the five-way distinction of gradience illustrated in Figure 1? Should the mid-value on a Likert-like scale (i.e. 3 or 4 depending on the adopted 5- or 7-point rating) correspond to slightly degraded or severely degraded judgments? As a matter of fact, both interpretations have been found in the corpus that constitutes the basis for this paper. Radford & Iwasaki (2015: 733) report that “[they] asked 20 experienced linguists [...] to judge the acceptability of the sentences [...], using a 5-point acceptability scale on which they were asked to rate each sentence as OK, or ?, or ??, or ?*, or *.” On the other hand, Lasnik (2014: 19) informs that in an experiment that he ran he “sent a list of 12 sentences to [his] department’s listserv (which mainly includes faculty, staff, graduate students, and undergraduate research assistants). Like the reviewer, [he] asked for the examples to be rated on a scale of 1 (completely well formed) to 5 (completely ill formed).” Unlike Radford & Iwasaki’s example, Lasnik’s formulation of experiment instructions plainly suggests that the middle of the scale must be closer to the ? level. The bottom line is that neither solution is perfect but allowing both of them to co-exist may produce results that are not comparable. If we follow Radford and Iwasaki (2015), then questionnaires may likely produce skewed results as respondents

tend to conceive of the mid-value as the neutral option (and degraded judgments are hardly neutral). However, the more common approach, as illustrated above from Lasnik (2014), assumes 3 on a 5-point Likert-scale to correspond to slightly degraded judgments and it raises the question of what judgments occupy the area between 1 and 3 (no dedicated notation available for ‘almost perfectly acceptable judgments’) on the one hand, and on the other hand, it forces degraded and severely degraded judgments to correspond to a single rating (number 4) on the scale.

2.2. Counterproductive creativity

Another aspect of the use of AJs that makes an easy target for criticism is the creativity observed in the number of various levels of acceptability. Despite the fact that the extended scale of acceptability consists of 5 distinct levels, as discussed in the previous section, linguistic articles are full of additional levels, which are typically very idiosyncratic and often unclear. A selection of such additional levels found in the studied papers is presented in (9–14).

- (9) ‡Who did you say to that I would buy the guitar?
(Bruening 2018: 390)
- (10) (??)The cocktail waitress entered the dining room who was wearing a blond wig.
(Reeve & Hicks 2017: 239)
- (11) (??)John didn’t make Mary NOT choose an expensive present, but he didn’t make her not NOT choose one either.
(Matsuoka 2013: 601, after Ritter & Rosen 1993: 539)
- (12) ??*If frankly he’s unable to cope, we’ll have to replace him.
(Haegeman 2010: 603)
- (13a) Fred denied that a certain boy talked to a certain girl.
(13b) ???I wish I could remember which boy to what girl.
(Lasnik 2014: 12)

- (14) ???/*The senator voted FOR the tax cut, but I don't know what AGAINST.

(Radford & Iwasaki 2015: 726)

It stands to reason that the notational chaos, which seems to be rampant among partially acceptable judgments, is not a welcome situation for linguistic papers. Naturally, the need for a finer-grained structure of partially acceptable judgments is perfectly understandable, but a 5-way acceptability gradience is subjective enough on its own, let alone an even more complex system. In situations where 5 levels are not enough, empirical methods should definitely be employed to validate the more subtle judgments.

One should also recognize other ways of showing the relative nature of acceptability, e.g. using mathematical notation, as in (15):

- (15a) This is the paper that we really need to find someone who understands.
 \geq
 (15b) This is the paper that we really need to find a linguist who understands.
 \geq
 (15c) This is the paper that we really need to find the linguist who understands.
 \geq
 (15d) This is the paper that we really need to find his advisor, who understands.
 \geq
 (15e) This is the paper that we really need to find John, who understands.
 (Hofmeister & Sag 2010: 372, after Kluender 1992)

2.3. Making informal AJs less informal

It must be acknowledged that in some papers attempts are made to make the data look less informal. Typically in such cases intuitions of additional informants are taken into consideration while determining the acceptability of an example. To illustrate such an attempt, let us quote from Bruening (2010: 45) who states that “I have checked most if not all of the data in this article with several other native speakers of English (between four and twelve) – some linguists, some not. The judgments appear to be quite robust. I have not found anyone who disagrees with the judgments reported here.” Similar descriptions

of small-scale acceptability judgment tasks are in fact commonplace (two more are described in the following section). The problem with such attempts is that they do not comply with any uniform methodology that is adopted by default in formal experiments, such as the choice of participants, a sample size, the selection of items, etc. That is why the results obtained in such ways cannot be meaningfully compared across different papers. Also, such results can be easily undermined (even if unfairly) on the ground of being biased (e.g. subjects may be aware of the arguments made by the author, etc.).

2.4. What should count as partial acceptability?

One more aspect of partial acceptability which is left at the discretion of each individual linguist is the decision whether to mark off any given example as partially acceptable or not. The following two examples should make this point more transparent.

In Yoshida et al. (2012: 485) the following three examples are discussed.

- (16a) Not every doctor₁'s knowledge of tax law or his₁ accountant's comments about the doctor's (knowledge) of medicine are useful.
- (16b) No semanticist₁'s opinion of syntacticians' attitudes toward morphology and his₁ students' opinion of phonologists' (attitudes) toward sociolinguistics were both expressed inappropriately.
- (17) [No parent]₁'s attitude toward politics should bias [DP his₁ children's religion].

As the possibility of co-variation is not straightforward, Yoshida et al. mention the small questionnaire in which “[they] consulted 10 native speakers of English regarding their judgment of [16a–b] and [17]. For [16a], 7 out of 10 native speakers accepted it and 3 did not. For [16b] and [17], 8 accepted the example and 2 did not.” The small sample notwithstanding, one can still see that in one case 30% of judgments pointed to the sentence being unacceptable and in the case of the two other sentences 20% of respondents rejected them. None of the sentences, however, has been marked as partially acceptable. It seems that at least a percentage mark could be used to signal their less than perfectly

acceptable status as otherwise other authors, especially non-native English speakers, may relate to these examples as fully acceptable.

To show that the previous example is not an isolated one, let us have a look at another case. In Matsuoka (2013: 594) the discussion concerns mental attitude (MA) adverbs and the three examples shown below are used.

(18a) John sent Bill *reluctantly* to the doctor. (John/Bill)

(18b) Mary put Susie *calmly* on the bed. (Mary/Susie)

(18c) Mary dragged her child *unwillingly* out of the bed. (Mary/her child)

Matsuoka explains that “[he] consulted twenty English speakers about the interpretation of MA adverbs in the sentences in [18] or similar ones. [...] Among the twenty speakers, fifteen answered that the adverbs in [18], or at least one of them, can be construed with either the subject or the object, whereas five replied that they can only be construed with the subject. The grammatical statuses given to the examples in the text represent those for tolerant speakers.” As before, the question is whether we should use the notation to indicate that an example is only acceptable to some speakers or is it enough to make a relevant comment under the example or in a footnote? Although in both cases the intentions of the authors are clearly communicated (the results of informal polls are fully disclosed), it seems that placing some notation before an example would be more appropriate as it eliminates the risk of the example ever being used as fully acceptable, especially, as noted before, by non-native speakers of English. Also, it must be remembered that such examples could be produced in much less studied languages where the process of filtering out questionable data works completely differently from English. Therefore, a simple solution and a good practice might be to put a percentage mark before such examples, which is done by some authors, as shown in (19), but does not seem to be a universal practice.

(19) %On this wall has never hung/will never hang a picture of Chomsky.
(accepted by some speakers)

(Bruening 2010: 60)

A percentage mark would effectively signal that the author has consulted the acceptability of an example, either informally or semi-informally, and the

example produces mixed or inconclusive responses. Anyone willing to cite such an example would instantly know that it must be treated with due caution or, preferably, properly tested using formal methods. For the time being, however, a percentage mark is only used sporadically and examples with a shaky status often find their way into papers unannotated. That is not to say that their authors want to smuggle in questionable examples into their research; most of the time their status is discussed in the paper. The fact remains, however, that no notation may still incorrectly suggest a fully acceptable status given that a range of notational symbols exist to mark partial acceptability.

3. The study of partial acceptability

In this study³ 40 articles published between 2010 and 2018 have been randomly selected from a previously assembled pool of 101 articles from *Linguistic Inquiry*, *Language*, *Syntax* and *Natural Language and Linguistic Theory*. The two main criteria have been employed to select articles:

- a) they had to contain English-only or predominantly English linguistic examples;
- b) they had to contain informal AJs or predominantly informal AJs.

The main aim of the study was to provide a comprehensive overview of the role of partial acceptability in generative papers, and, in particular, the study focused on:

- a) the description of ‘flavors’ or types of partially acceptable judgements, their distribution and the frequency of occurrence;
- b) the function of partially acceptable judgements (i.e. whether they have been used as a source of positive evidence / negative evidence or inconclusively).

³ This study is a follow-up to the study presented in Żychliński (2020), which reports on the use of partial acceptability in Polish. The conclusions reached in this paper are later compared with the results from Żychliński (2020).

While the first element of the analysis seems self-explanatory, the function of partial acceptability requires further clarification. One has to start with the fact that linguists who rely on informal judgments are sometimes accused of bias in the sense that their data are claimed to be tailored to specific needs of the theories they are developing. And so it happens quite frequently that critics single out specific examples and point to the fact that a given effect of acceptability is only achieved through a very careful selection of inflectional forms or lexical items, etc. However, a separate barrage of criticism comes from the fact that sometimes one may notice that partially acceptable examples are either used to support the points made by an author (in other words, they are treated on a par with fully acceptable judgments), or, conversely, they are used to show that a given construction cannot be used as positive evidence, as illustrated in (20) and (21), respectively, and discussed below.

(20) ?One of the students spoke to one of the professors about something,
but I don't know which to which about what.
(Lasnik 2014: 3)

(21) ?What_i do you regret [_{CP} that John ate t_i]?
(Sato 2012: 298)

(22a) You gave the last donut to who/?whom?

(22b) Who did you give the last donut to?

(22c) To whom/??who did you give the last donut?
(Sobin 2014: 405)

As for (20), we find that the author says that “[a]s far as I know, all of the published examples involve exactly two *wh*-phrases. However, three *wh*-phrases also seem possible” (Lasnik 2014: 3). This is a rather clear indication that the questionable (20) is treated as admissible, thus positive evidence. Example (21), on the other hand, is introduced with the statement informing that “[t]his analysis correctly accounts for the inability of *wh*-extraction out of a factive complement, as in [21], because the trace of the movement is not properly licensed (...)” (Sato 2012: 298). In this case, example (21) is explicitly taken as negative evidence. Apart from these two functions, the most

cumbersome group includes examples such as (22). For those examples, there is either no acknowledgement of their purpose, or their role is stated in uncertain terms. In the case of (22), we learn that “[i]n PPs involving *who/whom*, fronting the PP adds pressure to use *whom*, as in [22c]” (Sobin 2014: 405). What is adding pressure meant to suggest? Naturally, we understand that the sentence will improve upon using *whom* instead of *who*, but the status of (22c) remains unclear.

The subjective nature of the decision what to make of partial acceptability is precisely one of the main reasons why it has come under scrutiny in the present study. The idea is to see to what extent partial acceptability (which by its very nature implies that at least some speakers deem it more or less unacceptable) is used as either positive or negative evidence. The results will also demonstrate whether linguists prefer to err on the side of caution (and, thus, are more likely to reject a dubious judgment) or are more liberal in accepting shaky judgments as positive evidence.

In total, there have been 4491 acceptability judgments identified in the analyzed articles.⁴ The overall number is surprisingly high (more than twice the number of acceptability judgements found in the articles on Polish examined in Żychliński (2020) but the length of the papers is an important factor here – the Polish journals which were the main source of articles in the previous study largely contained shorter articles, averaging 20-30 pages, whereas the English journals used in the present study frequently contained considerably longer papers. Partially acceptable judgements accounted for 7% of all judgements, which amounted to 316 examples (see Figure 2). The following Sections 3.1 and 3.2 examine the types of partial acceptability and their function, Section 3.3 juxtaposes the results from English with the Polish results.

⁴ The identification of AJs is quite tricky. The obvious case is where an example has been invented by an author solely for the purpose of the paper and such examples constitute the bulk of the corpus of AJs discussed here. However, there is also a sizable group of examples that are cited from other sources, which means they are AJs made by others. In such cases they have also been counted as separate AJs on the uncontroversial assumption that the authors of the papers examined in the present study have accepted (unless stated otherwise) their acceptability rating. The same is true about much less frequent cases of examples which were taken from either corpora or authentic sources, following the same reasoning. Finally, the study has not covered papers which were entirely based on formally tested AJs, but the semi-informal ones (see Section 1.3) have also been included. What is perhaps most important, the criteria assumed in this study are identical to the criteria adopted in Żychliński (2020), which makes both studies comparable.

3.1. Types of partial acceptability

In 40 articles selected for the study, 35 (88%) contained partially acceptable judgments. Even though the exact number of ‘flavors’ or types of partial acceptability was 11 (? , ??, ???, ?*, ???/*, ??/*, ??*, ?‡, ‡, %, #), certain types were highly idiosyncratic (?‡ and ‡ have been used by only one author) and a few other types could be lumped together (no meaningful difference has been observed between ?*, ??/*, ??* and ???* or between ?? and ???). Thus, the picture which emerges presents us with 5 main types of partial acceptability. Table 1 shows each of them along with the frequency of their occurrence.

Table 1. Frequency of occurrence of partial acceptability in English.

Type of partial acceptability	Frequency of occurrence
?	21%
??	34%
?*	10%
%	2%
#	31%
Other	2%

The single most commonly found type of partial acceptability in English is the use of ?? (34%), which corresponds to degraded examples from Figure 1. The second most frequent notation is #,⁵ which accounts for 31% of all judgments. Slightly degraded examples are ranked third at 21% of all judgments, followed by severely degraded examples at 10%. There are only a few rather isolated instances of the use of the percentage mark⁶ and other notations.

⁵ The inclusion of the hash mark may be considered as questionable because the hash mark is typically used to describe semantic infelicity, and as such may be perceived as the semantic counterpart to an asterisk, which denotes ill-formedness (mostly of the grammatical type), thus not partial acceptability. The reason why the hash mark has been considered as a sub-type of partial acceptability is because there have been examples presented in Żychliński (2020) which show that a hash mark has also been used in a similar fashion to ?. It is important to point out that this use of # has not been observed in the English data.

⁶ A percentage mark, when it is used, typically refers to examples that are not uniformly acceptable (which often follows from a semi-informal test, e.g. an informal questionnaire among friends and colleagues or a small pool of respondents), as discussed in Section 2.4.

3.2. Function of partial acceptability in English

The crucial part of the study was the examination of the function of partial acceptability. Table 2 presents the distribution of functions of partially acceptable judgments.

Table 2. Distribution of function of partial acceptability.

Function	English
Positive evidence	20% (61 examples)
Negative evidence	71% (227 examples)
Unclear	9% (28 examples)

Partial acceptability in English seems to be treated with caution. Only 20% of all examples have been considered as a source of positive evidence while the bulk of partially acceptable examples (73%) are used as negative evidence and the smallest group of examples (9%) comprises those without a clear indication of their function. Table 3 shows the distribution of the functions across different types of partial acceptability.

Table 3. Distribution of functions of partial acceptability across types.

Type of partial acceptability	Positive evidence	Negative evidence	Unclear
?	16%	2%	3%
??	2%	28%	4%
?*		10%	
%	2%		
#		31%	
Other			2%

When it comes to the specific distribution of partial acceptability, it becomes apparent that slightly degraded judgments are predominantly taken as positive

evidence (or treated on a par with acceptable judgments), whereas degraded and severely degraded judgments are largely used as negative evidence. Judgments without specified function can be mostly found among slightly degraded and degraded examples. The hash marked judgments constitute a sizable portion of all examples and in English, where they are used solely to indicate semantic infelicity, they are uniformly used as negative evidence.

3.3. The use of partial acceptability in English and Polish – a comparison

As stated before, Żychliński (2020) studied the use of partial acceptability in the linguistic research on Polish.⁷ The results of that study, which the present study has been modeled on, will now be compared with the English results.

First of all, let us look at the general composition of acceptability judgments in English and Polish. In 1855 acceptability judgments in Polish, there were 128 occurrences (7%) of partial acceptability. This shows a striking similarity between Polish and English in the extent to which partial acceptability is used, despite a considerably higher number of data points in English, which is illustrated in Figure 2.

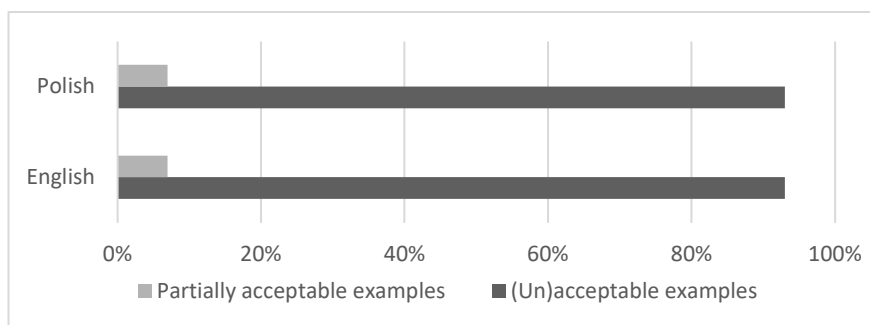


Figure 2. The structure of acceptability judgments in English and Polish.

⁷ The forty articles comprising the Polish corpus were mainly taken from *Poznań Studies in Contemporary Linguistics*, *Studies in Polish Linguistics*, *Studia Anglica Posnaniensia* and *Roczniki Humanistyczne*. Additional sources (e.g. authors' websites, post-conference publications or LingBuzz) were necessary as the primary sources were insufficient.

Let us have a look next at the frequency of occurrence of corresponding types of partial acceptability in English and Polish in Table 4.

Table 4. The frequency of occurrence of partial acceptability in English and Polish.

Type of partial acceptability	Frequency in English	Frequency in Polish
?	21%	44%
??	34%	21%
?*	10%	19%
%	2%	5%
#	31%	11%
Other	2%	n/a

The most revealing contrast is seen in the use of slightly degraded judgments. In the Polish data this type of partial acceptability has been identified more than twice as frequently as in the English data. Degraded examples, on the other hand, account for 34% of all examples in English compared with 21% for Polish and severely degraded examples are more frequent in Polish (19%) than English (10%). The use of the percentage mark is very limited in both languages and the English data contained considerably more semantically infelicitous hash marked examples.

The function of partial acceptability will be compared next. Table 5 presents the comparison of the functions of partial acceptability in English and Polish.

Table 5. Function of partial acceptability in English and Polish.

Function	English	Polish
Positive evidence	20% (61 examples)	46% (59 examples)
Negative evidence	71% (227 examples)	31% (39 examples)
Unclear	9% (28 examples)	23% (30 examples)

Even though partially acceptable examples in English and Polish account for roughly the same chunk of all examples, Table 5 shows that they serve different purposes in each of the two languages. Only 20% of partially acceptable examples are used as positive evidence in English, as opposed to almost half (46%) of similar examples in Polish. At the same time, the majority of English partially acceptable examples (71%) are used as negative evidence, as opposed to only 31% of such examples in Polish. Finally, examples without specified function constitute almost a quarter of all examples in Polish, whereas in English they only account for 9% of all examples.

Last but not least, let us consider the similarities and differences in the distribution of functions of partial acceptability across all types for English and Polish, as presented in Table 6.

Table 6. Distribution of function of partial acceptability across types in English and Polish.

Type of partial acceptability	Positive evidence		Negative evidence		Unclear	
	English	Polish	English	Polish	English	Polish
?	16%	39%	2%	5%	3%	1%
??	2%	2%	28%	6%	4%	13%
?*			10%	13%		5%
%	2%	3%				2%
#		2%	31%	7%		2%
Other					2%	

The most significant differences in the treatment of partial acceptability in English and Polish have been highlighted in Table 6. First of all, the Polish partially acceptable data contains a large proportion of slightly degraded judgments (39%) which are used as positive evidence, compared to 16% of such examples in English. At the same time, the English data shows a more extensive use of degraded judgments (28%), which are largely used as negative evidence. Moreover, one can see that in the case of both slightly degraded and degraded judgments there is some variation as to the function which is assigned to them (especially in Polish 5% of ? examples are used negatively as opposed to 2% in English). The proportion of severely degraded judgments is

similar in both languages and they are uniformly used as negative evidence. The table also reveals a comparable treatment of the percentage mark and a degree of inconsistency in the use of the hash mark. The latter notation is only used to mark examples used as negative evidence in English, but in Polish it is sometimes used in a parallel fashion to a question mark. One more notable difference is the proportion of unclear examples, i.e. those whose function is not conclusively stated in the text. These examples occur three times as often in Polish as in English.

To summarize, the study of partial acceptability in English and Polish shows that despite an equal representation of partially acceptable judgments in both languages, the way they are used is not identical. In Polish, slightly degraded AJs are used more often and they are mostly used as positive evidence, whereas in English, the largest group contains degraded judgments, which are used as negative evidence. Also, judgments without specified function are far more frequent in Polish, which is not a welcome finding as a risk arises that they may be bent to suit specific needs, if only by non-native speakers of Polish. One possible explanation for the discrepancies between the two languages may be the fact that there is far less research focused on Polish as opposed to English and, thus, more examples pass as positive evidence as they are not so readily filtered out by fellow linguists. This would also explain why the majority of partially acceptable examples in English serve as negative evidence – more discussions and peer reviews (not only formal, but during conferences, invited talks, etc.) may result in certain judgments being moved to the fully acceptable category, while questionable data may end up in the degraded category. Also, linguists working on English may prefer to err on the side of caution (and be less willing to accept a questionable judgment as positive evidence) as they are more likely to anticipate that the data in circulation will still be subject to scrutiny from a large field of linguists fluent in English. Finally, Western journals may follow more stringent rules regarding the use of partially acceptable judgments, although it is only a speculation as there is no direct evidence for that.

4. Conclusion

This paper has been concerned with a subset of acceptability judgments, namely partially acceptable judgments. Even though partially acceptable judgments are widely used in linguistic literature, there is very little, if any,

literature that offers a meta-linguistic description of their application. It is hoped that a comprehensive discussion of the issues related to the use of partial acceptability followed by a study of the role that partial acceptability plays in English and Polish linguistic papers will not only shed more light on this slightly obscure category but also provide some guidelines on how to improve the quality of informal judgments. Dispensing with ad hoc levels of acceptability or ensuring that any partially acceptable judgment is clearly defined in terms of its function are just two examples of good practices that should be considered by authors.

5. Sources of data

- Bruening, Benjamin. 2010. Language-particular syntactic rules and constraints: English locative inversion and *do*-support. *Language* 86(1). 43–84.
- Bruening, Benjamin. 2018. CPs move rightward, not leftward. *Syntax* 21(4). 362–401.
- Gehrke, Berit & Louise McNally. 2015. Distributional modification: The case of frequency adjectives. *Language* 91(4). 837–870.
- Haegeman, Liliane. 2010. The movement derivation of conditional clauses. *Linguistic Inquiry* 41(4). 595–621.
- Hofmeister, Philip & Ivan A. Sag. 2010. Cognitive constraints and island effects. *Language* 86(2). 366–415.
- Iatridou, Sabine & Ivy Sichel. 2011. Negative DPs, A-movement and scope diminishment. *Linguistic Inquiry* 42(4). 595–629.
- Kluender, Robert. 1992. Deriving islands constraints from principles of predication. In Helen Goodluck & Michael Rochemont (eds.), *Island constraints: Theory, acquisition and processing*, 223–258. Dordrecht: Kluwer.
- Kubota, Yusuke & Robert Levine. 2016. The syntax-semantics interface of ‘respective’ predication: a unified analysis in Hybrid Type-Logical Categorical Grammar. *Natural Language and Linguistic Theory* 34. 911–973.
- Lasnik, Howard. 2014. Multiple sluicing in English? *Syntax* 17(1). 1–20.
- Matsuoka, Mikinari. 2013. On the notion of subject for subject-oriented adverbs. *Language* 89(3). 586–618.
- Moulton, Keir. 2013. Not moving clauses: Connectivity in clausal arguments. *Syntax* 16(3). 250–291.
- Postal, Paul M. 2004. *Skeptical linguistic essays*. Oxford: Oxford University Press.
- Postal, Paul M. & Geoffrey K. Pullum. 1982. The contraction debate. *Linguistic Inquiry* 13(1). 122–138.
- Radford, Andrew & Eiichi Iwasaki. 2015. On swiping in English. *Natural Language and Linguistic Theory* 33. 703–744.
- Reeve, Matthew & Glyn Hicks. 2017. Adjunct extraposition: Base generation or movement. *Syntax* 20(3). 215–248.

- Ritter, Elizabeth & Sara Thomas Rosen. 1993. Deriving causation. *Natural Language and Linguistic Theory* 11. 519–555.
- Sato, Yosuke. 2012. Multiple Spell-Out and contraction at the syntax-phonology interface. *Syntax* 15(3). 287–314.
- Sheehan, Michelle L. 2013. Some implications of a copy theory of labeling. *Syntax* 16(4). 362–396.
- Sobin, Nicholas. 2014. Th/Ex, agreement, and Case in expletive sentences. *Syntax* 2014(4). 385–416.
- Yoshida, Masaya. 2010. “Antecedent-contained” sluicing. *Linguistic Inquiry* 41(2). 348–356.
- Yoshida, Masaya, Honglei Wang & David Potter. 2012. Remarks on “Gapping” in DP. *Linguistic Inquiry* 43(3). 475–494.

References

- Cowart, Wayne. 1997. *Experimental syntax: Applying objective methods to sentence judgments*. Thousands Oaks, CA: Sage.
- Culicover, Peter W. & Ray Jackendoff. 2010. Quantitative methods alone are not enough: Response to Gibson and Fedorenko. *Trends in Cognitive Sciences* 14(6). 234–235.
- Dąbrowska, Ewa. 2010. Naive vs. expert intuitions: An empirical study of acceptability judgments. *The Linguistic Review* 27. 1–23.
- Gibson, Edward & Evelina Fedorenko. 2010. Weak quantitative standards in linguistic research. *Trends in Cognitive Sciences* 14(6). 233–234.
- Phillips, Colin. 2009. Should we impeach armchair linguists? *Japanese/Korean Linguistics* 17. 49–64.
- Schütze, Carson T. 1996. *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Chicago: University of Chicago Press.
- Schütze, Carson T. 2020. Acceptability ratings cannot be taken at face value. In Samuel Schindler, Anna Drożdżowicz, & Karen Bröcker (eds.), *Linguistic intuitions: Evidence and method*, 189–214. Oxford: OUP.
- Sprouse, Jon & Diogo Almeida. 2012. Assessing the reliability of textbook data in syntax: Adger’s Core Syntax. *Journal of Linguistics* 48(3). 609–652.
- Sprouse, Jon & Diogo Almeida. 2017. Design sensitivity and statistical power in acceptability judgment experiments. *Glossa: A Journal of General Linguistics* 2(1). 1–32.
- Sprouse, Jon, Carson T. Schütze & Diogo Almeida. 2013. A comparison of informal and formal acceptability judgments using a random sample from Linguistic Inquiry 2001–2010. *Lingua* 134. 219–248.
- Wasow, Thomas & Jennifer Arnold. 2005. Intuitions in linguistic argumentation. *Lingua* 115. 1481–1496.

Żychliński, Sylwiusz. 2020. The role of reduced acceptability in Polish linguistic articles. In Karolina Drabikowska & Anna Prażmowska (eds.), 197–210. *Exploring variation in linguistic patterns*. Lublin: Wydawnictwo KUL.

Corresponding author:

Sylwiusz Żychliński
Faculty of English
Adam Mickiewicz University
Collegium Heliodori Święcicki
Grunwaldzka 6
60-780 Poznań
Poland
sylwiusz@amu.edu.pl