Anomalies and Exceptions

Hubert Haider, Dept. of Linguistics, Univ. Salzburg (2006)

1. Anomalies and exceptions: Anomalies in the observed data patterns are usually construed as exceptions in the grammar of the data patterns. ‘Anomaly’ is a characterization of data properties in terms of ‘normalization’ expectations. ‘Exception’ is the reconstruction of an anomaly in terms of a rule with a restriction. Whether an anomaly is best characterized as the local effect of an exceptional rule or as a global system effect (result of conflicting but otherwise exceptionless rules) is an empirical issue. Not every anomaly is an exception, though. It may be a mere processing difficulty (see below).

The conceptual difference between anomaly judgements sampled from performance data and exceptionality ascriptions to rules of grammar must not be obscured by equivocation. The source of the performance data is a self-evaluation of the mental processing of a stimulus, the second one, namely the grammar ascription, is an attempt to model the grammar-related properties of the performance data. Note that the grammar-related aspects are only a subsystem of the complex cognitive computations whose composite output is the global acceptability judgement for a given stimulus.

Anomalies may be the result of processing difficulties in the absence of any exceptional trait in the grammar of the given construction. Strong garden path effects are the best examples:

(1) a. Man glaubt, dass Max Musiker vorgestellt bekamen¹
    one believes that Max musicians introduced got_{3rd Pl.}
    b. Das sind es²
    this are it

The perceived anomaly of (1a) is the difficulty of identifying scrambling in the absence of overt case markings. If you replace ‘Max’ by ‘den_{acc} unbekannten Max’ (the unknown Max), and (optionally) ‘Musiker’ by ‘viele Musiker’ (many musicians), the anomaly disappears. In (1b), both pronouns are singular, but the finite copula is marked for plural. This mismatch is perceived as an anomaly. Only when the informant is pushed to realize that the predicate ‘es’ (it) may as a predicate refer to a plural entity, the judgement changes from deviant to perfect.

An exception is a restriction on the range of a rule to a subset of its possible range of application (2b). In (2b), x^n is exempted from the range of the universally quantified rule predicate P.

(2) a. ∀ P(x)
    b. ∀[\{X-x^n\}] P(x) exception: the range is the set X minus the member x^n

There is no denying that grammars are exceptional to a certain extent. Even if exceptions exist on the level of individual items (3), the frequent case is a restriction on a subclass (4). Here is an example for a restriction on the level of an individual lexeme. In every Germanic language, the cognate of ‘genug’ (enough) is anomalous, since it follows rather than precedes the modified item:

¹ Scrambling of an object without overt case across the subject in the German get-passive variant produces strong deviance feelings for informants. Thanks to M. Schlesewsky for the datum.
² Thanks to M. Bierwisch for reporting me this datum (source: E. Lang). Informants reject it first, but they accept it immediately once you ask for instance “Is this a possible answer for: Are this really 47 envelopes?”
(3) a. gross genug, big enough, groot genoeg (Dutch), ...

An example for a subclass restriction is the restriction on the German ‘Ersatzinfinitiv’ construction\(^3\) that is itself triggered by the exceptional avoidance of the participle form for verbs that select a bare infinitive (modal, perception, causative verbs). In German, but not in Dutch, the construction is restricted to the finite clause. It is ungrammatical in infinitival clauses (4b) in German, but not in Dutch (4c).

(4) a. dass er es hat essen wollen
that he it has eat want

b.*ohne es zu haben essen wollen
without it to have eat want

c. zonder het te hebben willen eten
without it to have want eat

Binding by pronouns in German provides examples of an anomaly caused by inconsistent grammar requirements: on the one hand, an antecedent must c-command the bindee (5a),\(^4\) on the other hand, a dative pronoun must follow an accusative pronoun in German. So, the German restriction on pronoun order rules out binding of an accusative reflexive by a dative pronoun, since the dative would have to precede (in order to fulfil binding) and thereby it would violate the order restriction embodied in the Grammar of German pronouns (5b). This leaves (5c) as the only option and rules out (5b), in accordance with Featherstone’s experimental findings, reported in his figure 1.\(^5\)

(5) a.*Wir haben sich\(^i\)\(_{\text{Acc}}\) (selbst) ihnen\(^i\)\(_{\text{Dat}}\) überlassen
we have them (to) them\(_{\text{Refl}}\) (selves) left

b.*Wir haben ihnen\(^i\)\(_{\text{Dat}}\) sich\(^i\)\(_{\text{Acc}}\) (selbst) überlassen

c. Wir haben sie\(^i\)\(_{\text{Acc}}\) sich\(^i\)\(_{\text{Dat}}\) (selbst) überlassen
we have them (to) them\(_{\text{Refl}}\) (selves) left

However, this account does not fully cover the Dative-Accusative anomaly with respect to binding in German. Although a dative may be a binder for an anaphor (6a), it appears to be disqualified if the bindee is a co-argument (6b) of the dative binder. In addition, there is another anomaly involved. A dative not only does not bind in these cases, it also interferes with binding between the subject and the object (6d), if it intervenes.

(6) a. Er hat den Leuten\(^i\)\(_{\text{Dat}}\) über einander/sich\(^i\) erzählt / Biografien von einander/sich\(^i\) gezeigt

b.*Er hat den Leuten\(^i\)\(_{\text{Dat}}\) einander/sich\(_{\text{Acc}}\)\(^i\) vorgestellt

c. Er hat die Leute\(^i\)\(_{\text{Acc}}\) einander/sich\(_{\text{Dat}}\)\(^i\) vorgestellt

d.??Wir\(^i\) haben diesem Mann einander/uns\(^i\) vorgestellt

e. Wir\(^i\) haben einander/uns\(^i\) diesem Mann vorgestellt

\(^1\) Descriptively speaking, the (finite) auxiliary that would trigger the participial form on the preceding verb is fronted and the would-be-participle appears as the bare infinitive form.

\(^2\) The only apparent exception is binding by the nominative subject: ‘Hat sich\(_{\text{Refl}}\) jemand\(^i\) geirrt?’ But in this case, there is an agreeing item, i.e. the finite verb, that c-commands (see Frey 1993).

\(^3\) Note that the subjects in the experiment seem to treat binding by objects as an undefined case, since they allow personal pronouns (Principle B) on a par with reflexives (Principle A) in this case (see Featherstone’s figure 1, ‘ndp’ and ‘nar’). This would follow if the core case of binding in German is binding between the subject and an object and binding between objects is un-/ill-defined in the application of German grammar by the informants.
These data have not been part of Featherstone’s experiment, but they are crucial for the dative anomaly. In my view, the anomaly is not yet fully understood. We do not yet see clearly enough whether it involves a genuine exception or is just a result of conflicting interactions of otherwise unexceptional rules of binding.

2. Discreteness or gradience: Data judgements are necessarily gradient, grammars are discrete. Judgements are the composite, unconscious result of the interaction of various components (grammar, information structure, stylistic preferences, implicit comparison with potential paraphrase variants, ease of parsing and interpretation, anticipative judgements of the experimenter’s expectations, ...). Even if each of these components would produce a discrete evaluation value, the aggregate would be a gradient function.

Well-formedness as defined by human grammars is discrete, that is, a matter of yes or no. There is no such thing as ‘75% grammatical’. Crucially, well-formedness as a property determined by grammar, must not be equated with the introspective attribution of the quality of well-formedness by informants. Their judgements are reports on introspection experiences and these are surely not discrete. From a theoretical point of view, grammatical well-formedness is not adequately characterized as a matter of cumulated (relative) weights. The fact that informants asked for grammaticality judgements are at a loss in certain cases does not prove that grammaticality is gradient. Discreteness shows in the majority of cases if ‘clear cases’ are tested. To make an issue testable is difficult, of course. It is a matter of a meticulous test design, including pilot studies, and it requires ultimately also a lot of experience, giftedness and even luck on the experimenters’ side.

There is no need to succumb to experimentingly testing the robustness of the contrast between (7a) and (7b). It is evident. The head of the attribute phrase must be adjacent to the head-initial phrase it modifies (see Haider 2004a, on the ‘edge effect’). Second, an adverbial modifier precedes the modified element, with the exception of ‘genug’ (see (3), above). Because of the adjacency restriction, (7b) is deviant. Some speakers resort to (7d), but they never use (7e). Is the difference between (7b), (7d) and (7e) a gradient one? Yes, it is in terms of acceptability, but it is not in terms of well-formedness. (7b) and (7d) are equally ungrammatical. Why is (7d) considered less deviant? It is felt to be the best solution in a no-win situation: If you stick to the order required by the idiosyncrasy of ‘genug’, you would have to inflect and thereby turn it into a kind of fake head of the attribute, otherwise it violates the adjacency requirement (7d). What remains is the violation of inflecting an uninflectable item that is not the head. In sum, gradience is often the result of dealing with a situation in which two requirements are in conflict.

(7a) ein genügend deutliches Beispiel
   a sufficiently clear example

---

6 Müller (1995, sect. 4.5) suggests that the Dative is base generated lower than the accusative and then raised, and therefore it could not serve as a binder for the accusative. Immediate counterevidence for this claim is the fact that i) Dutch does not allow object scrambling but has Dative < Accusative as the obligatory order, and ii) that a raised argument ought to be a possible binder (see: The men' seem to each' other to be incompetent).

7 Optimality theory suggests discrete rules with a relative weighting in terms of their violability ranking. Until now, the proponents have not produced a universal theory of weighting, however. Without a UG of ranking, a weighted system is not learnable: If the perceived input deviates from the child’s interim grammar output, the necessary changes in the interim grammar involve intractable computations of comparing alternative rankings
b. *ein deutliches genug Beispiel
   a clear enough example

c. *ein deutliches genügend Beispiel

d. ??ein deutlich genuges Beispiel

e. *ein deutlich genügendes Beispiel (*, in the reading of (7a))

It is an undeniable fact that the results gained in psycholinguistic experiments are heavily influenced by the design of the experiment. If the experimenter was in the lucky situation to be able to control most of the potentially intervening variables, the results will be close to the ideal of being representative for the issue under experimentation, but in most cases this is hard to achieve.

Featherstone’s experiment on superiority is a good case. The experimental data document a preference for the patterns congruent with superiority in German as well as in English, but for German, the preference is characterized as not ‘categorical’ in the sense that “speakers would not choose to use structures violating” the superiority constraint. This characterization does not pay enough attention to intervening variables, however. The experiment tests only a subset of superiority cases, namely direct questions. Indirect questions (8c,d) would have been the better choice. The contrast between English and German would become even clearer.

(8) a. It is unclear what belongs to whom
   b.*It is unclear to whom what belongs
   c. Es ist unklar, was wem gehört
       it is unclear what whom\text{Dat} belongs
   d. Es ist unklar, wem was gehört
   e. Wem gehört was?
   b.*To whom does what belong?

Direct questions presuppose a context since they presuppose that a possible answer exists and the answer is an appropriate choice of discourse participants for the questioned elements. Second, for this choice, the order of the wh-elements provides the sorting key. For instance, in (8e), the sorting key is ‘wem’ (to whom). So, the elements of a subset of the set of possessors are mapped on elements of the set over which ‘was’ (what) ranges, namely the set of possessed elements. Indirect questions do not require an answer, so they do not need a discrete choice of potential discourse participants, and so the question of the sorting key is not relevant, and therefore the information structure effect of the order of wh-elements is less salient.

What the experimental data confirm is this. If an informant is presented a sentence in isolation, (s)he implicitly embeds it in a potential discourse situation and judges the information structure. If you have to choose between (9a) and (9b), you will easily identify the order in (9a) as congruent with the base order. The preference of (9a) is a preference of the contextually unmarked information structure.

(9) a. Wer hat was zur Party mitgebracht?
   b. Was hat wer zur Party mitgebracht
   c. Ich möchte gar nicht wissen, was wer zur Party mitgebracht hat
   d.*I do not want to know what who has brought along to the party
What a simple comparison between English and German fails to honour is a cross-linguistic generalization. Superiority does not only hold for English, it holds for any VO language (Haider 2004), but not for OV languages. The absence of superiority is not just a property of German, it is a property of OV languages in general, as for instance, Japanese (10a,b).

(10)a. Nani o dare ga katta no
    what-OBJ who-SUB bought Q-PRT - ‘What did who buy?’

  b. Dare ga naze kita no
    who-SUB why came Q-PRT - ‘*Who came why?’

  c. Wer hat wie/weshalb/wann/wo protestiert?
    who has how/why/when/where protested

  d. *Who protested when/where/*why/*how? 

The reason is this (see Haider 2004): In VO, the VP-internal subject is preverbal and therefore not in the directionality domain of the verb. In OV, any argument, and in particular the subject, too, is in the directionality domain of the verbal head. So, the subject in VO, but not in OV, needs a functional head as a directional licenser. This is the grammar theoretic source of the existence of obligatory functional subject positions in VO and the peculiar behaviour of an in-situ wh-subject (see Haider 2004, 2005, for details).

The that-trace effect is just a facet of this phenomenon, but an ill-understood one. First, in English, the effect is absent if an adverbial intervenes (Browning 1996). Second, neither German nor any other OV language punishes that-trace-structures. Third, languages differ with respect to the general transparency of C-introduced clauses. In German, speakers of northern varieties object to any wh-extraction out of dass-clauses whereas Southerners are free extractors. This has been documented already by Paul (1919: 321f.). He devoted a subsection of his German Grammar to long distance wh-dependencies and referred to them justly as ‘Satzverschlingung’ (sentence intertwining). In his collection, he documents plenty of cases of that-t-violations for interrogative and relative clauses. Andersson & Kvam (1984) tested extractions out of that-clauses in various locations in Germany and not only showed a contrast between extractors and non-extractors, but also a kind of adaption effect for non-extractors that tolerate extractions by others.

So, that-t-violations need to be checked as carefully as Torris (1984) did in her dissertation. She showed that systematic extractors for wh-constructions are systematic extractors for the other cases (comparatives, relative clauses, long distance topicalization), too. On the other hand, there are extraction admitters and they show an unsystematic behaviour. They tolerate extractions by others but long distance extraction out of dass-clauses is not part of their grammar. This is reminiscent of Andersen’s (1973) concept of ‘via-rules’ for communities in which the common vernacular language is fed by microparametrically contrasting grammars.

In sum, subject extraction out of a C-introduced finite clause is on the one hand subject to constraints on extraction out of C-introduced finite clauses in general, and on the other hand, it is restricted by constraints for traces in functional spec position. The former restriction applies to variants of German, the latter one does not.

---

i) Who do you think that *(under these circumstances) would disagree?
Just like experimental data, corpora do not speak for themselves. They require evaluation and interpretation. We have learnt that water freezes at 0° centigrade. If you want to test it, you will find out that the results are gradient and that you may be puzzled by exceptions and learn about the influence of intervening third factors, like impurities, pressure fluctuations, etc. Corpora are performance records, and performance is unavoidably prone to influences of third factors, including simple mistakes and imperfections. A strict methodology of corpora evaluation is still wanting. Nevertheless, corpora may serve heuristic purposes. If something is a robust phenomenon, a big enough corpus will reflect this. More subtle relations are hard to immediately assess by mere corpora inspection.

3. Methods and theories: The ideal grammar is exceptionless (Featherstone, sect. 1). But, grammars of human languages are not ideal. Unlike platonic objects (e.g. logical calculi), they are biologically grounded, cognitive systems for a culturally formed dynamic behaviour, namely human languages. What we perceive as exceptions are compromises in the fine-tuning of a complex modular system. Some of them are externally geared (diachronic relics), some of them internally (inconsistent rule demands).

From the methodological point of view, it is a crucial question as to whether a perceived anomaly is the reflex of an exceptional trait of the system or just apparent. It is apparent if what we perceive as exception is the result of inadequately modelling a phenomenon that would turn out as regular in a wider context of an adequate account. This is exception by error or, in other words, a scientist’s deficiency. A main business of science is to remove these ‘exceptions’ by testing and changing their theories.

How can we distinguish between ‘real’ exceptions and apparent exceptions? Featherstone’s claims rest on the interpretation of the results of a refined method of gaining introspection data from informants (‘magnitude estimation’). This method in my opinion is not reliable and valid enough to call into question generalizations arrived at by a systematic comparative study of grammar by the scientific community of experts.

First, naive introspection is notoriously erratic – across individuals and across categories for a single individual – except for the most robust contrasts. Second, an informant’s judgement is an aggregate of all factors that influence a ‘this is how I myself would (not) say’-judgement. So, third, informants’ judgements would have to be accompanied by a protocol of what the informants report as the crucial traits of the stimulus that (s)he has based the judgement on. Fourth, the patterns of reaction have to be tested for consistency (across the categories under examination) and for retest stability. Fifth, the minimal battery of statistic analysis tools needs to be employed in order to guarantee that the sample is representative, that the correlations are significant, that the results are solid enough to stand up against the null hypothesis, and so on. If linguists apply psychological or sociological methods, they are bound to comply with the required methodological standards developed for these methods.

In sum, in the face of the results and interpretations of Featherstone’s investigations, I do not feel compelled to give up my conviction that grammars determine discrete characteristic functions (‘well-formed’ vs. ‘ill-formed’) for linguistic expressions. What appears to be gradient is not the grammar, but the reactions of the test subjects. Models that employ weighted rules (‘violation costs’) necessarily obscure this important difference: discrete systems produce
gradient outputs if the output is mediated by additional interacting system. This is the case for human languages. Grammar theory models a cognitive capacity for a discrete symbol management algorithm. Grammar theory does not model the cognitive architecture of language production and perception. This is the realm of processing theories. Performance data bear only in a highly indirect way on competence issues.

References
Browning, Margaret A. 1996. CP-recursion and ‘that-t’-Effects. Linguistic Inquiry 27:237-255
Featherston, Sam. (this vol.) Three types of exceptions – and all of them rule-based.