Grammar of Statements Involving "Chance"

Milo Schield[1], Tom Burnham[2]
[1]Augsburg College, Minneapolis, MN
[2]Cognitive Consulting, San Antonio, TX

Abstract

Statements involving "chance" are commonly used by statistical educators in presenting probability, confidence levels and statistical significance. The chance family of grammars includes these keywords: "chance," "odds," "risk," "likelihood" and "probability." This paper identifies instances of statements involving "chance" that are syntactically ambiguous in distinguishing between the uncertain outcome (the variable element) and the generating process (the fixed or conditional element). The syntax of a large number of statements involving "chance" is examined to see if the collocates help in making this distinction. Numerous conjectures on syntax are offered that may help in making this distinction. This paper concludes that chance grammar is extremely ambiguous. Much more work is needed to make stronger claims about the role of the syntax and the design of a computer program to decode chance grammar.

1. Problems with Statements Involving "Chance"

Kahneman, Slovic and Tversky (1982) found that people regularly made logical mistakes in dealing with probability and likelihood.

Kahneman et al told their subjects that "Bill is 34 years old. He is intelligent but unimaginative, compulsive and generally lifeless. He was strong in mathematics but weak in social studies and humanities." Respondents were asked to rank order the following statements by their probability: (A) Bill is an accountant, (J) Bill plays jazz for a hobby, (A & J) Bill *is* an accountant *who* plays jazz for a hobby. Kahneman et al argued that this description of Bill made "A" much more likely than "J" and that the probability of the conjunction (A & J) should be no more than that of either (A) or (J) and thus no more than that of the least likely: "J". Overall, 85% of respondents ranked the likelihood of the conjunction (A & J) as being between that of "A" and "J". The conjunction effect mistake was made by 92% of statistically naïve subjects, 83% of statistically sophisticated subjected and 86% of intermediate subjects.

Subjects in the intermediate group were asked to consider the argument that "the probability that Bill is both an accountant and a jazz player cannot exceed the probability that he is a jazz player, because every member of the former group is also a member of the latter." More than two-thirds of them "agreed, after some reflection, that their answer was wrong because it was at variance

with the conjunction rule." See p. 92-95.[1] Consider some similar cases.

Case #1: In which are the odds greater in a future year[2]:
1a. that a massive flood occurs somewhere in America in which thousands die.
1b. that an earthquake in California causes a massive flood in which thousands die.

Case #2: "Which is more likely on a given day?"[3]
2a. that a woman will give birth
2b. that a 20 year-old woman will give birth.

Case #3: Which is the greater risk in a future year?
3a. that a US dam will break and kill a 100 people.
3b. that a broken US dam will kill a 100 people.

In each case, it seems that Kahneman et al would argue that (a) is more likely than (b) since specifying a conjunction of two conditions makes (b) more restrictive than (a). Kahneman et al named the failure to see (b) as the correct answer the "conjunction effect" (p. 91). It is also known as the "conjunction fallacy:"[4] the failure to see that the probability of a conjunction is never more than the probability of either component since $P(A \cap B) \leq P(A)$ and $P(A \cap B) \leq P(B)$. Their explanation for such mistakes was psychological: chance is subtle and our minds aren't designed for subtlety.

But perhaps there is an alternate explanation. One possibility is that people don't have a good idea of any of the chances involved, so thinking about chance involving multiple factors can produce cognitive overload.

One way to eliminate the possibility of respondent misunderstanding is to formulate a question in which the probabilities are clearly known to all the respondents.

---

[1] Kahneman et al (p. 92) asked respondents to rank the following statements by their probability: a) Linda is a bank teller or b) Linda is a bank teller and is active in the feminist movement. While the grammar of (b) clearly states an intersection, readers may be substituting "who" for "and" or else treating the "and" as an "or."

[2] http://members.fortunecity.com/templarser/wotlogic.html

[3] www.rci.rutgers.edu/~gbc/RPDM/heuristics.htm

[4] http://en.wikipedia.org/wiki/Conjunction_fallacy

Case #4: The lead author asked statistical educators:

"When flipping a fair coin, what is the chance that *a head on the 1st flip will be followed by a head on the 2nd?*" Choose from these three answers.

4a. ½   4b. ¼   4c. Don't know or can't tell

Based on the lead author's estimated count of the hands shown at an ASA meeting[5], two-thirds chose 4a and a fifth chose 4b so about a tenth chose 4c. Based on the lead author's estimated count of the hands shown at an IASE meeting[6], slightly over half chose 4a and only two chose 4b so nearly half chose 4c.[7] By the reasoning of Kahneman et al, most statistical educators are guilty of the conjunction effect mistake.

But are they? Statistical educators certainly understand the probabilities in flipping a fair coin twice. The issue is not the subtlety of chance. The issue is grammar.

There are two interpretations for this question.

4a   The subject (head on 1st flip) is given and the predicate is uncertain, so $P(H2 \mid H1) = ½$.

4b   The entire clause is uncertain, so $P(H1 \& H2) = ¼$.

The two answers depend on how one interprets the question: as conditional (4a) or as an intersection (4b).

Case #4 is the same as #1-3. In each case, (b) involved a subject-predicate clause where readers could interpret the uncertain element as just the predicate (conditional statement) or as the entire clause (conjunction). Interpreting (b) as conditional would entail choosing (a).

Look back at the probability statements involving Bill. In the combined statement both elements appear after the verb "is", so this is not a case where readers identify the uncertain element as the predicate. Look instead at the relative pronoun "who." Recall that in percentage grammar, "the percentage of W who are P" presents W as the whole or the given and P as the part or the variable element. Perhaps the respondents interpreted the statement "Bill is an accountant who plays jazz for a hobby" as a conditional statement: P(Bill plays jazz for a hobby | Bill is an accountant). If being an accountant (A) and playing jazz for a hobby (J) are positively associated, then the conditional probability, $P(J|A)$, can be greater than the smaller unconditional probability: $P(J)$.

The probability of the conjunction of two outcomes is always less than (or equal to) the probability of their conditional relationship.[8]

---

[5] 80 attendees at the 2007 ASA JSM session on Statistical Literacy.

[6] 40 attendees at the lead author's talk at the 2007 IASE in Portugal.

[7] This may be due to the higher percentage of attendees who learned English as a second language.

[8] $P(A\&B) = P(B|A)P(A)$ so $P(A\&B) \leq P(B|A)$ and $P(A\&B) \leq P(A|B)$

Notice that these cases of ambiguity all involve "that": "the probability that" (in the case of Bill the accountant), "the odds that" (in case #1 of people dying in floods), "the likelihood that" (in case #2 of women giving birth), "the risk that" (in case #3 of a broken dam causing a flood) and "the chance that" (in case #4 of two flips of a coin.) In each of these cases, "that" introduced a clause having two syntactic elements either or both of which could be the uncertain element.

It is interesting to note that in the first 127 page of Kahneman et al (1982) where they discussed people's misunderstanding of chance as a measure of uncertainty, that 60 of the 82 uses of chance-family grammar involve a chance family keyword (e.g., "chance," "odds," "risk," "likelihood" or "probability") followed by "that" where "that" introduces a subject-predicate clause. Perhaps something about such clauses facilitates ambiguity.

But rather than focus just on clauses following "chance that," we begin by examining the syntax of all statements involving "chance" as a noun.

2.   Chance Family of Grammar

Chance, likely, prevalent, risk, odds, likelihood and probability – along with accuracy and confidence – are concepts that point to the existence, identity or location of some uncertainty and can measure the amount of this uncertainty. The amount of uncertainty can be indicated qualitatively (e.g., "Gamblers have a *small chance* of winning consistently.") or quantitatively (e.g., "the chance of getting two heads in the next two flips of a fair coin is 25%.")

Schield (2000) concluded that syntax of statements involving "chance" was similar to that involving "risk," "odds," "likelihood" and "probability" so they should be grouped into a single family called the "chance family" since "chance" is the most common member. The syntax of "likely" and "prevalent" is slightly different so they were located in a separate family.

Schield (2000) found that the prevalences for the members of the chance family as a percentage of their total number were "chance" (25%), "odds" (20%), "risk" (15%), "likelihood" (10%) and "probability" (5%). Since that time, certain other usages have been identified as fitting this grammatical structure. These include words such as "confidence" (e.g., "a 20% confidence that/of") and "accuracy" (e.g., "a 20% accuracy"). These latter uses are very specialized so they are not discussed in this paper.

3.   Chance Grammar in Everyday Use

This analysis is based on all the lines (11,674) in the Harper-Collins Cobuild Bank of English database in fall, 1999 that contained "chance" as a noun. The goal

is to identify the syntax that distinguishes the uncertain element from the generating process: the process having an outcome that is uncertain.

Appendix A presents the 100 most common words that appeared *immediately after* "chance." Syntactically, the types of words (and their percentage of total uses) are prepositions, infinitive and subordinate/relative clauses (82%), conjunctions (3.9%), pronouns (3.4%), verbs (2.3%), nouns (0.9%), articles (0.2%) and other (1.5%) and unknown (5.4%). "Other" includes known words that do not fit any of the preceding groups.

The four most common words immediately following "chance" – and their relative prevalence – are: "to" (41%), "of" (27%), "for" (3.6%) and "that" (2.9%). While "of" and "for" are prepositions, "to" introduces the infinitive verb and "that" in this case is something very different.[10] It seems that chance grammar spawns all kinds of odd-ball children.

When "chance" is followed by pronouns, nouns or articles (e.g., "chance he won"; "chance China won"; "chance the team won"), there is often – but not always[9] – an unstated "that" following "chance." If the virtual uses (3.4%) of "that" are added to the actual uses (2.9%), then the top four words and their prevalence are "chance to" (41%), "chance of" (27%), "chance that" (6.3%) and "chance for" (3.6%).

Appendix A also presents the 100 most common words that appeared *immediately before* "chance." Syntactically, the types of words (and the percentage of their uses) are articles (48%), adjectives (31%), pronouns (7%), prepositions (2%), verbs (0.3%), other (0.8%) and unknown (11%). The two most common words preceding "chance" – and their prevalence – are two articles: "a" and "A" (26%) and "the" and "The" (22%).

Conjecture #1. "Chance" functions as a measure of uncertainty when "chance" is followed by a description of the uncertain element. It appears that this occurs when "chance" is a noun preceded by an article, pronoun or adjective (e.g., "some chance of rain,": "a 50-50 chance of winning") and followed by "that," a preposition, infinitive or by an article, pronoun or noun.

Conjecture #2. It appears that "chance" refers to a concept when (1) "chance" is a noun followed by a verb or conjunction (e.g., "chance is subtle"; "chance and uncertainty are two sides of the same coin"), (2) "chance" is preceded by a preposition or verb. Lines involving chance as an explanation may be indicated when "chance" is in a prepositional phrase such as "by chance," "of chance" (e.g., "by means of chance"), and "to chance" (e.g., "to be due to chance" or "addicted to chance"). (3) "chance" is adjective modifying a follow-

ing noun or pronoun (e.g., "a chance meeting").

4. "Chance To"

Appendix B presents the 100 most common words following "chance to" and their prevalence.

All of the words immediately following "chance to" are verbs since "to" introduces the infinitive form.

The eight most common follow-on words after "chance to" are win (14%), get (6%), see (6%), make, play, do, go and be.

The follow-on after "chance to" may involve a single word (e.g., "chance to win"), a clause with an external subject (e.g., "chance to get the part", "chance to see a game," "chance to be a winner"), a clause an internal subject (e.g., "chance to see our team win") or combination thereof.

The infinitive verb may involve a copula. While other verbs can involve a single word predicate (e.g., "to win"), copulas require follow-on words (e.g., "to see a play," "to be a winner").

Conjecture #3. It appears that when "chance to" is followed by a single word – or by a simple phrase containing an article and a single word – before a sentence break (period or comma) or a predicate, then that single word is the uncertain element or outcome.

Conjecture #4. It appears that if "chance to" is followed by a verb introducing either a complex phrase involving at least two grammatical items that are not articles (e.g., "the chance to win the Preakness twice"), multiple phrases (e.g., "chance to win the Preakness after winning the Kentucky Derby," "chance to roll a head after rolling a head") or a clause with a subject and predicate (e.g., "chance to see our horse win the Preakness") then the uncertain element cannot be determined from the syntax alone. The analysis of clauses is undertaken in the "chance that" section.

The words immediately preceding "chance to" can be classified into four syntactic categories as articles (73%), pronouns (9.3%), adjectives (14%), other (0.3%) and unclassified (4.4%).

Conjecture #5. It appears that none of the top-100 words preceding "chance to" support the use of "chance" as referring to a concept.

5. "Chance Of"

Appendix C presents the 100 most common words following "chance of" and their prevalence. These follow-on words can be classified into five syntactic categories: words ending in "ing" (48%), articles (14%), nouns (8%), pronouns (5%), Other (2%) and unclassified (13%).

---

[9] "one of those chance affairs" "Thanks to a chance break"

The follow-on after "chance of" may involve a single word (e.g., "chance of winning"), a phrase (e.g., "chance of being a winner," "chance of a win"), a clause (e.g., "chance of Jill winning," "chance of seeing our team win") or combination thereof.

Conjecture #6. It appears that if "chance of" is followed by a single word phrase (e.g., "chance of winning") or an article and a single word (e.g., "chance of a win"), then that single word indicates the uncertain element.

Conjecture #7. It appears that if "chance of" is followed by a single multi-word phrase (e.g., "chance of a winning season"), then the entire phrase indicates the uncertain element.

A possible exception to this rule involves the use of sports grammar. In sports grammar the uncertain outcome and the population or generating process are both indicated in an adjective-noun phrase where the whole is a natural whole. Consider the chance of competed pass in football. This could mean the chance of a completion (the uncertain outcome) given that the play involves a pass (the chance that a pass is completed) or the chance of a completed pass as the uncertain outcome from all the plays in a game (the chance that a play results in a completed pass). Note that this could be restated as the chance of a pass completion. Thus, the uncertainty could be the entire phrase or the adjective or noun in an adjective-noun phrase (e.g., "chance of a male smoker.") See Schield (2000).

Conjecture #8. When "chance of" is followed by multiple phrases, the syntax is not always sufficient to determine the variable element (e.g., "the chance of rolling a second head after the first head" or "the chance of a flood following an earthquake").

A discussion of subject-predicate clauses following "chance" ("chance of" or "chance that") is presented in the next section.

6. "Chance That" Statements

Appendix D presents the 100 most common words following "chance that" and their prevalence.

The phrase, "chance that,"[10] is always followed by a clause. E.g., "chance that our team will win is small."

Conjecture #9. Identifying the uncertain element in "chance that" clauses depends on the type and voice of the verb.

- When the verb has an active voice and the predicate is an event or changeable condition, the uncertain item is typically the predicate. E.g., "The chance that the batter hits a home run."

- When the verb has a passive voice and the subject is an event or changeable condition, the uncertain item is typically the subject. E.g., "The chance that a home run is hit by the batter."

- When the verb is intransitive and the subject and predicate are unchangeable conditions, the uncertain item is typically the entire clause. E.g., "the chance that the man is Asian." This usage may reflect an application of percentage grammar where "that" always indicates the part. However, the percentage "that" is a relative pronoun, while the chance "that" is something different.

Conjecture #10. The ambiguity in clauses following "chance that" may apply to many clauses following "chance to" and "chance of." E.g., "the chance to see (of seeing) the team win" or "the chance to see (of seeing) a win by our team."

7. "Chance For" Statements

Appendix E presents the most common words following "chance for" can be classified into four syntactic categories: articles (33%), pronouns (30%), nouns (15%), other (3%) and unclassified (19%).

In some cases "chance for" is equivalent to a "chance to" statement. "The chance to win by our team" is equivalent to "the chance for a win by our team" or "the chance for our team to win."

8. Analysis

The foregoing analysis of collocates appears of rather limited value in identifying syntax that reliably distinguishes the uncertain element from the generating process in "chance" statements. The following conjectures are based on the authors' experience with "chance" statements.

Conjecture #11. The population being sampled or the process generating an uncertain outcome is

- always indicated by "if" (a preposition), "given" (a past participle) and "assuming" (a present participle) regardless of where they are placed relative to the keyword "chance."

- always introduced or qualified by relative pronouns such as "when," "while" and "where."

Conjecture #12. The uncertain element or outcome always follows (never precedes) the keyword "chance."

---

[10] In "chance that," "that" is a conjunction joining two noun clauses. In this usage, "that" does not appear as a relative pronoun (e.g., "the road that we took"). In this usage, "that" may be a combination of a "reporting that" (e.g., "The judge said that the accused is guilty) and a nominalizer "that" (e.g., "That the accused was guilty was obvious.") Professional grammarians differ on these points and we make no claim of being professional grammarians.

9. Indicating the Uncertain Element

Chance grammar can be very succinct (e.g., "the chance of rain"), which may be why it is so popular. Being succinct may involve omitting – or imprecisely designating – the information necessary to distinguish the uncertain element from the generating process. It may be that other actions are necessary to make this distinction.

Conjecture #13. To insure unambiguous statements involving "chance," place the generating process before the keyword "chance" so only the variable element and any optional relative pronouns follow "chance." (E.g., "Having gotten one head on the first flip and then flipping a fair coin the second time, what is the chance that it [the second flip] will be a head?" Alternatively in a more Frequentist form, "In a sequence of pairs of unbiased flips of a fair coin in which the first flip is a head, what is the chance of heads on the second flip?"

Conjecture #14. When the keyword "random" or "randomly" is used, it will typically describe the generating process (e.g., "the chance that the sample mean of a randomly selected sample will be more than 2 standard deviations from the population mean is less than 5%" or "the chance that two randomly flipped coins will both be heads is 25%").

Conjecture #15. Ambiguity between the generating process and the uncertain outcome within a clause can be eliminated by including the word "random" or "randomly" appropriately and using the future tense to indicate the uncertain element.

See how this can eliminate ambiguity. It converts "When flipping a fair coin, w*hat is the chance that a head on the 1ˢᵗ flip will be followed by a head on the 2ⁿᵈ?*" into "When flipping a fair coin, *what is the chance that a head on the first flip followed by a second random flip will give a head?*" or "When flipping a fair coin, w*hat is the chance that two successive random flips will give two heads?*"

It converts "The chance that a 95% confidence interval will include the population parameter" into "The chance that a random sample will generate a 95% confidence interval containing the population parameter."

10. Chance Grammar-Checker Design

One goal of this analysis is to design a program that can decode user-generated descriptions and comparisons of ratios using ordinary English that involves keywords such as "chance." See Burnham and Schield (2005).

It was hoped that this analysis would generate design recommendations or constraints. Given the lack of firm conclusions, it seems premature to set guidelines to help the program distinguish the uncertain outcome from the generating process, or to distinguish when

"chance" is used to designate an idea or concept from when "chance" is used to measure uncertainty

While there may be cases where a general-purpose program can make these distinctions reliably, these may be so limited in number or so severely restricted in form as to be of little use.

11. Future Chance-Related Projects

One project would be to test the conjectures contained in this paper (or variations thereof). This may entail (a) cleaning the data to exclude "chance" at the end of a sentence when the following period is missing, (b) including the virtual members of "chance that" along with the explicit members to increase the number of usages analyzed, (c) analyzing the syntax involving the other members of the chance family: risk, odds, likelihood and probability, (d) analyzing other sources such as those written by professional educators or statisticians, and – most importantly – (e) analyzing the meaning and/or ambiguity generated by having multiple phrases and/or clauses following "chance." One goal would be to identify those grammatical forms that are less likely to be ambiguous based solely on their syntax and to see what fraction of ambiguous statements can be restated without becoming overly cumbersome.

A second project would be to evaluate the cognitive status of chance statements – the semantics that distinguish objective and subjective probability. This requires an analysis of the conceptual differences between the formal, axiomatic approach of mathematical probability, the "objective" empirical approach of Von Mises (1981) and the pseudo-Frequentist approach (using chance to describe any relative frequency).

A third project would be to investigate the inferential status of chance-related statements. One inference involves going from past to future. Given a certain percentage, prevalence or incidence of events in the past, under what conditions are we warranted in describing this as a chance or probability? Another inference involves going from a group to a single individual. Can an objective Frequentist statement of chance be meaningfully applied to a single individual?

12. Conclusion

Statements involving "chance" can be ambiguous. More work is needed to identify those features of chance grammar that can reliably distinguish the uncertain element from the generating process. These statements involving "chance" are similar to the statements involving "probability" that are commonly made by statistical educators. Statistical educators should be aware of these ambiguities and develop a professional opinion on how one interprets potentially ambiguous statements involving "chance."

Appendix A. Chance Statement Syntax

This analysis is based on all the text that contained "chance" classified as a noun in the Harper-Collins Cobuild Bank of English database[11] as of fall, 1999. This extraction generated 11,674 lines. Lines having a length of 120 characters[12] were used for this analysis. The collocates (adjacent words) were analyzed using Monoconc.[13] Data was saved as a txt file, opened in Word, transferred to Excel and then back to Word.[14]

Selecting[15] on "chance" with a space before and after – generated 10,082 instances. The 100 most common words[16] *immediately following* "chance" can be grouped into seven syntactic categories:

- Prepositions, subordinate/relative clauses and infinitives (8,296): to, of, for, that, in, at, after, with, on, when, against, if, they, from, by, TO, before, over, off, To, out, FOR.
- Conjunctions (392): and, as, but, because, than, or, so, though ["though" may be an adverb], then.

- Pronouns (345): he, I, you, it, we, this, there, which, whatsoever, here, she, you've, I'll, something, you'll, it's, he'll, I've.
- Verbs (232): is, came, was, would, had, have, has, do, comes, did, will, offered, fell, went, does, take, are.
- Nouns (86): meeting, encounter, Saloon, remark, meetings, discovery, encounters, conversation, events.
- Articles (39): the, a.
- Other real words (149): just, two, no, going, quite.

The 100 most common words *immediately preceding* "chance" – sorted by relative frequency[17] – can be classified into six syntactic categories:

- Articles (4,857): a, the, The, A.
- Adjectives (3,090): no, good, any, best, last, little, better, another, great, every, more, real, only, second, cent, first, much, one, big, fair, less, percent, No, outside, greater, 50-50, Any, Last, realistic, reasonable, excellent, slim, rare, same, final, early, golden, fighting, slightest, strong, Second, some, unique, glorious, decent, even, high, possible, equal, 50, obvious, slight, fine, easy, greatest, genuine, ideal, outstanding, higher, simple, scoring, gilt-edged.
- Pronouns (688): your, his, their, my, this, what, Your, her, What, our, its, YOUR.
- Prepositions (230): by, of, after, to, that, By, on.
- Verbs (7): was.
- Other (537): and, cracking, or, each-way, as, Mm, 17, Maine, medal,

The words preceding chance – sorted by mutual information[18] instead of by frequency – are quite different.

---

[11] www.collins.co.uk/books.aspx?group=153

[12] A longer line is more likely it is to include multiple instances of chance so Monoconc generates duplicates. Use file 07L120sr1n.txt.

[13] Collocates were generated using Monoconc 2.0. See Athelstan at www.athel.com/. Procedure: (1) File/LoadCorpusFile: give file name. (2) Concordance/Search: enter search pattern; Select Options; Increase Maximum Search Hits from 10,000 to 12,000. (3) Select Concordance / SaveAsFile: give file name for selected lines. (4) Select Frequency / CollocateFrequencyData. (5) Select Frequency / SaveAsFile: give file name for concordance data. Note that L1 and L2, give the collocates and their number of uses to the left of the search term one or two words; R1 and R2 give those to the right. The Monoconc defaults show those collocates with three or more uses.

[14] To transfer saved concordance frequency data to Excel, open the saved txt data in Word and copy/paste the data into Excel. To convert the Excel cell data into a continuous stream of Word data, create =WordCell&" ("&#cell&"), " in Excel. Copy this formatted data and paste it into Word. In Word, convert the cells from Table to Text and then use Replace to eliminate all paragraph codes (^p).

[15] To eliminate "chance" preceded or followed by a non-space, the advanced search with regular expression "\schance\s" was used with "Ignore case of letters". This selection generated 10,082 lines of output. A subsequent control code could have been eliminated using "\schance\s[a-z]" but this eliminated the first letter of the collocate.

[16] The 100 most common words immediately following "chance" – sorted by frequency – are: to (4,114), of (2,684), for (365), that (294), in (218), and (205), at (147), after (75), with (74), on (70), he (64), when (55), I (52), as (50), is (49), you (48), but (44), meeting (40), against (37), if (36), came (34), the (31), was (30), it (30), we (29), because (27), they (26), than (24), or (22), this (21), from (21), by (21), now (20), would (18), again (17), had (16), TO (16), before (14), there (14), which (14), have (14), whatsoever (13), has (13), do (12), so (12), she (11), just (10), here (10), encounter (10), alone (9), today (9), No (9), a (8), you've (8), over (7), off (7), only (7), I'll (7), comes (7), something (6), all (6), did (6), Saloon (6), will (6), To (6), remark (6), meetings (6), out (5), discovery (5), you'll (5), well (5), it's (5), encounters (5), like (5), offered (5), fell (5), went (5), next (4), does (4), p (4), conversation (4), though (4), quarter (4), yet (4), he'll (4), actually (4), I've (4), events (4), away (4), take (4), FOR (4), soon (4), two (4), no (4), then (4), tomorrow (4), are (4), NS (4), going (4), quite (4). Of these 10,082 lines, 70% involve the top three words immediately following "chance," 80% involve the top eight, 90% involve the top 34 and 95% involve the top 100.

[17] The 100 most common words immediately preceding "chance" – sorted by relative frequency – are: a (2,600), the (2,123), no (372), your (265), good (236), any (204), best (182), last (181), little (159), better (155), another (148), his (143), great (133), every (109), more (107), **by** (87), real (75), only (67), their (65), The (65), second (63), A (63), cent (56), first (51), much (51), one (51), big (49), **of** (44), my (41), fair (36), less (35), percent (33), this (32), what (30), Your (29), No (28), outside (28), after (28), greater (27), 50-50 (27), **to** (26), that (25), Any (25), Last (23), realistic (22), her (21), What (21), reasonable (20), excellent (19), slim (18), rare (17), same (16), our (16), its (16), final (15), and (15), early (14), golden (14), fighting (13), By (13), slightest (13), strong (12), Second (12), some (11), unique (11), get (11), glorious (11), decent (11), even (11), high (10), possible (10), equal (9), YOUR (9), cracking (9), or (9), had (9), each-way (8), 50 (8), as (8), creating (8), obvious (8), slight (7), fine (7), Mm (7), on (7), easy (7), greatest (7), genuine (7), was (7), ideal (7), outstanding (7), hell's (7), 17 (7), higher (7), Maine (6), fantastic (6), simple (6), medal (6), scoring (6), gilt-edged (6). Of these 10,082 lines, 45% involve the top two words immediately preceding "chance," 50% involve the top three, 60% involve the top 8, 70% involve the top 15 and 89% involve the top 100.

[18] The 36 most common words immediately preceding "chance" – sorted by mutual information – are earthly, remotest, phantom, Clifford, slightest, treble, realistic, slim, cracking, birdie, Maine, scant, glorious, gasp, decent, sporting, reasonable, edged, slender, sheer,

Appendix B: "Chance To" Syntax

Selecting on "chance to" generated 4,012 uses. The words immediately following "chance to" are always verbs since "to" introduces the infinitive form. The 100 most common words immediately following "chance to" – sorted by frequency[19] – can be classified as copulas[20] or others. Being a copula[21] is highly contextual.

The words preceding "chance to" shown below[22] can be classified into four syntactic categories:

- articles (2,989): the, a, A, The.
- pronouns (383): your, their, his, Your, this, YOUR, my, her, our, My, its
- adjectives (566): last, another, great, no, first, big, second, better, good, rare, little, one, only, best, fi-

nal, unique, any, much, more, every, golden, early, real, fair, get, ideal, Second, glorious, perfect, excellent, Last, once-in-a-lifetime, 50-50, greater, EXTRA, simple, less, late, gilt-edged, same, exclusive, decent, equal.

- other (14): get, had, gets.

Appendix C: "Chance Of" Syntax

Selecting on "chance of" generated 2,681 uses. The 100 most common words immediately after "chance of" – sorted by frequency[23] – can be grouped into five syntactic categories:

- Words ending in "ing."[24] (1,274): winning, getting, being, making, playing, going, becoming, finding, staying, beating, surviving, having, seeing, avoiding, doing, taking, catching, reaching, qualifying, developing, landing, pulling, living, breaking, achieving, bringing, dying, finishing, succeeding, improving, saving, working, gaining, defeating, turning, stopping, picking, pipping, securing, adding, remaining, starting, scoring, keeping, resolving, clinching, competing, coming, building, persuading, meeting, holding, carrying, hitting, completing, returning, passing, putting, ending, joining, recovering, changing, progressing, regaining.
- Articles (367): a, the, an.
- Nouns (210): success, survival, victory, life, peace, promotion, bass[25], recovery, Europe, rain, conger[25], bream[25], happiness, sole[25].
- Pronouns (134): that, him, me, it, them, your, this, you, his, their, us, something, anyone.
- Other (51): any, ever, another, one, actually, good.

---

fair, rare, better, medal, slight, best, scoring, smallest, golden, another, real, outstanding, greater, great, fighting and little.

[19] The 100 most common words immediately following "chance to" – sorted by frequency – are: win (406), get (164), see (158), make (111), play (107), do (88), go (83), be (77), prove (75), put (68), have (65), talk (58), show (58), take (52), meet (44), buy (38), join (37), look (34), try (33), catch (32), work (30), say (29), become (29), develop (26), save (25), vote (24), enjoy (23), learn (22), help (21), watch (19), impress (19), give (19), speak (18), tell (18), escape (17), share (17), experience (17), find (17), live (16), hear (16), gain (16), start (16), test (16), build (16), score (15), explore (15), stay (14), visit (14), come (14), change (13), turn (13), set (13), follow (13), discuss (13), run (13), move (12), sample (12), fly (12), bring (12), pick (12), grow (12), rest (11), break (11), study (11), think (11), shine (10), compete (10), relax (10), use (10), recover (10), end (10), taste (10), tour (9), strike (9), improve (9), read (9), reach (9), establish (9), participate (9), ask (9), enter (9), grab (9), return (8), create (8), demonstrate (8), discover (8), succeed (8), add (8), WIN (8), clinch (8), settle (8), spend (8), travel (8), drive (8), avoid (7), own (7), open (7), sit (7), appear (7), Learn (7). Of the 4,012 uses, 20% involve the top four words following "chance to," 40% involve the top 16 and 69% involve the top 100.

[20] The list of link verbs depends on whether the verb is followed by an adjective, a noun group or a prepositional phrase. See Collins Cobuild (1990). Per 3.134, verbs that can be used as link verbs with the adjective as the complement include: (1) be; (2) appear, feel, look, prove, seem, smell, sound, taste; (3) become, come, fall, get, go, grow, keep, remain, stay and turn. Some of the verbs in the second group are often followed by an infinitive (e.g., she appears or seems to be happy). The verbs in the third group involve changing or staying the same. Per 3.193, verbs that can be used as link verbs when a noun group is the complement include (1) be, become, remain; (2) feel, look, prove, seem, sound; (3) constitute, make, represent; (4) comprise, form. Per 3.175, link verbs that are used with a prepositional phrase as the complement include appear, fall, feel, get, keep, remain, seem and stay – in addition to "be" per 3.174.

[21] A combined list of copulas includes 23 verbs: appear, be, become, come, comprise, constitute, fall, feel, form, get, go, keep, look, make, prove, remain, represent, seem, smell, sound, stay, taste and turn.

[22] The words with at least three uses preceding "chance to" – sorted by frequency – are: the (1,542), a (1,386), your (242), last (103), another (69), A (38), great (35), no (35), their (31), his (30), first (27), Your (27), big (21), second (21), better (18), The (18), good (17), rare (14), this (14), little (13), one (13), only (12), best (11), final (11), unique (10), any (10), much (10), more (10), every (10), golden (9), YOUR (9), my (9), her (8), early (8), real (8), fair (7), get (7), ideal (6), Second (6), THE (5), our (5), glorious (4), perfect (4), My (4), had (4), its (4), excellent (4), Last (4), once-in-a-lifetime (3), 50-50 (3), greater (3), EXTRA (3), simple (3), less (3), late (3), gilt-edged (3), gets (3), same (3), exclusive (3), decent (3), equal (3).

[23] The 100 most common words immediately following "chance of" – sorted by frequency – are: winning (295), a (263), getting (122), being (111), making (84), the (72), success (57), survival (54), playing (54), going (39), becoming (38), finding (33), an (32), staying (26), beating (25), that (25), surviving (23), having (23), seeing (22), him (21), avoiding (21), doing (20), taking (18), me (16), catching (16), reaching (15), any (15), qualifying (15), victory (15), it (13), developing (13), life (13), landing (12), them (12), peace (12), promotion (11), bass (11), pulling (10), living (10), ever (10), another (10), breaking (9), achieving (9), bringing (8), your (8), dying (8), finishing (8), recovery (7), succeeding (7), improving (7), saving (7), working (7), gaining (7), Europe (7), this (7), one (7), you (7), defeating (7), turning (6), stopping (6), his (6), picking (6), rain (6), pipping (6), their (6), securing (6), adding (6), remaining (5), starting (5), conger (5), scoring (5), keeping (5), resolving (5), clinching (5), competing (5), coming (5), actually (5), building (5), persuading (5), meeting (5), holding (5), us (5), carrying (5), something (4), bream (4), hitting (4), completing (4), returning (4), good (4), passing (4), putting (4), ending (4), happiness (4), joining (4), recovering (4), changing (4), anyone (4), sole (4), progressing (4), regaining (4). Of these 2,681 uses, 20% involve the top two words immediately following "chance for," 40% involve the top nine and 76% involve the top 100.

[24] Words ending in "ing" include gerunds (verbs used as nouns and ending in "ing") and verbs (present participles). "Wining" is a gerund in "winning is fun" but a present-participle verb in "the team is winning" or "the chance of winning the game."

[25] "Chance of [catching some] bass, conger, bream or sole."

The words preceding "chance of" are shown below.[26]

Appendix D: "Chance That" Syntax

Selecting on "chance that" generated 306 uses. The words immediately following "chance that" – sorted by frequency[27] – can be grouped into four categories:

- Pronouns (117): it, he, they, you, she, I, we, he'll, this, she's, your, you'll.
- Articles (59): the, a.
- Verbs (9): was[28], is[29]
- Prepositions[30] (6): if[31], in[32].
- Nouns (13): Denmark, Clive, fate, Ebayseen.
- Other (3): there[33].

When chance is followed by a verb, chance is an idea that is being described – not a measure of uncertainty.

When chance is measuring uncertainty, it seems that the follow-on after "chance that" *must* involve at least one clause (e.g., "chance that it will snow," "chance that a coin is fair").

The words immediately preceding "chance that"[34] include articles (a, the), adjectives (good, any, every, no, little, real, slight, fair, excellent, reasonable, slightest, percent, 50-50, high, happy) and prepositions (by, of). The prepositions "by"[35] and "of"[36] typically use chance as an idea rather than as a measure of uncertainty.

Appendix E: "Chance For" Syntax

Selecting on "chance for" generates 301 uses. The most common words immediately following "chance for" – sorted by frequency[37] – can be classified into four syntactic categories:

- Articles (99): the, a, an.
- Pronouns (87): that me, him, it, you, them, us, her, everyone, his, those.
- Nouns (46): peace, people, Israel, survival.
- Other (8): lasting, all.

In some cases "chance for" is a variation of a "chance to" statement. "The chance to win by our team" becomes "the chance for a win by our team" or "the chance for our team to win."

The words preceding "chance for" are shown below.[38]

REFERENCES

Burnham, Tom and Milo Schield (2005). "An Online Grammar-parsing Program That Decodes Ordinary English Descriptions and Comparisons of Percentages and Rates." *ASA Proceedings of the Section on Statistical Education.* [CD-ROM] P. 2202-2210. See www.StatLit.org/pdf/2005BurnhamSchieldASA.pdf.

Collins Cobuild. *English Grammar.* Harper Collins Publishers, 1990.

Kahneman, Daniel, Paul Slovic and Amos Tversky, Editors (1982). *Judgment under uncertainty: Heuristics and biases.* Cambridge University Press.

Schield, Milo (2000). Statistical Literacy: Difficulties in Describing and Comparing Rates and Percentages. *ASA Proceedings of the Section on Statistical Education.* P. 176. See www.StatLit.org/pdf/2000SchieldASA.pdf.

Von Mises, Richard (1981). *Probability, Statistics and Truth.* Dover Press.

[26] The words with at least three uses preceding "chance of" – sorted by frequency – are: a (478), the (383), no (239), any (136), good (113), little (110), better (102), best (98), more (71), great (56), cent (48), real (39), much (33), every (32), The (28), percent (26), only (25), realistic (23), greater (23), less (22), his (21), outside (20), 50-50 (19), your (18), their (18), last (17), fair (17), Any (16), reasonable (14), fighting (13), slim (10), excellent (10), some (8), same (8), one (7), genuine (7), hell's (7), strong (7), higher (7), all (6), outstanding (6), our (6), NO (5), big (5), cracking (5), high (5), fantastic (5), increased (5), No (5), even (5), equal (5), fine (5), another (4), 50 (4), greatest (4), slightest (4), four (4), possible (4), Slim (3), 17 (3), three (3), woman's (3), 0 (3), first (3), million (3), earthly (3), small (3), decent (3), smallest (3), fat (3), second (3), remote (3).

[27] The words with at least three uses immediately following "chance that" – sorted by frequency – are: the (51), it (24), he (18), they (13), you (12), she (11), we (11), I (9), a (8), was (5), he'll (5), this (5), you'll (5), she's (4), Denmark (4), is (4), Clive (3), there (3), fate (3), your (3), Edbaysaan [horse](3), if (3), in (3).

[28] "took every chance that was available."

[29] "This is a chance that is offered to few." "It cannot be mere chance that is responsible for our dismissal."

[30] These are interpolated adverbial phrases/clauses which could be moved elsewhere. Regular chance grammar continues after the interpolation. These interpolations are often enclosed between commas.

[31] "A chance that if you keep smoking you will die."

[32] "It was largely by chance that in 1985 she took an Observer."

[33] "There is an 84% chance that there will be an oil spill."

[34] The words with three or more uses before "chance that" – sorted by frequency – are: a (73), the (35), good (25), every (16), any (16), by (14), little (9), no (9), real (5), reasonable (4), slight (4), of (4), slightest (4), excellent (4), fair (4), percent (3), 50-50 (3), high (3).

[35] E.g., "It's not by chance that;" "it was quite by chance that;" "it's just by chance that;" "it was only/simply/hardly by chance that"

[36] "There are layers (was an element) of chance that;" "knew from sitting in the still eye of chance that fate was not random"

[37] The words with at least three uses immediately following "chance for" – sorted by frequency – are: the (47), a (46), me (22), peace (19), him (19), you (16), people (15), Israel (8), them (7), us (7), an (6), her (5), everyone (4), survival (4), lasting (4), all (4), his (4), those (3).

[38] The words with at least three uses immediately preceding "chance for" – sorted by frequency – are: a (148), the (29), another (24), good (23), last (22), great (15), best (13), little (7), no (5), A (5), The (4), big (4), LAST (3), its (3), her (3), only (3)