Granularity shifting: Experimental evidence from numerals vs. adjectives

Theoretical prerequisites. Simple/short expressions like the numeral *ten* tend to have coarser & more approximate interpretations than complex/long ones like *nine point three* which tend to have fine & precise interpretations (Krifka 2002, 2007). Krifka (2007) models this phenomenon via an interval [n – n/s, … , n + n/s] representing a normal distribution with mean *n* (the precise interpretation) and standard deviation *s* (the level of approximation). With smaller deviation values *s*, the interpretation gets more precise. Assuming equal probabilities for scale points, it follows that approximate interpretations are more probable than precise ones and interpretations relative to coarse scales are more probable than fine ones. Assuming a preference for short expressions, the shortest among any set of numerals with (roughly) the same distribution is associated with the most probable (coarse & approximate) interpretation, while the longer ones must be interpreted more precisely.

According to Lewis (1979), a shift from default to finer granularity and/or precision level is a natural discourse move, but the opposite shift is not. For example, we may state that the Netherlands is flat, presupposing default coarse granularity g, and then point out that it is actually a bit bumpy by shifting to a finer criterion g_p, taking as evidence bumps we previously ignored. However, we cannot state that the Netherlands is bumpy and smoothly proceed to say that it is actually flat, ignoring bumps that we previously regarded as relevant.

Hence, the simple numeral *10* in (1a) is expected to have a default coarse and approximate interpretation 'about ten' ($[\text{ten}]_g$ = [9.5,10.5)), and *9.5* is expected to have a precise interpretation ($[9.5]_{gp}$ = [9.5]). Since the latter is part of the former ($9.5 \in [9.5,10.5)$), in (1a) B's utterance is understood as confirming A's utterance. By contrast, we expect that following an utterance of a complex numeral (e.g., *9.5* in (1b)), the interpretation of a simple numeral (e.g., *10*) would be finer and more precise than the default ($[\text{ten}]_{gp}$ = [10]). Since $10 \neq 9.5$, and since numerals tend toward upper-bounded or 'exact' readings (e.g. *9.5* and *10* are understood as conveying 'exactly 9.3' and 'exactly 10', in particular when interpreted along precise criteria), in (1b) B's utterance is understood as contradicting A's. Therefore, we predict higher agreement rates in coarse-to-fine contexts like (1a) than fine-to-coarse ones.

1) a. A: *Stock Exchange fell by ten percent.* B: *Yes (#No), it fell by nine point five percent*.
   b. A: *Stock Exchange fell by nine point five percent.* B: *No (#Yes), it fell by ten percent*.

Adjectival modification as in *slightly dirty* and *completely clean* triggers a granularity shift too, rendering, e.g., hardly visible dirt specks that are by default ignorable relevant in judging cleanliness. However, the non-default fine & precise interpretation of *dirty* following *slightly dirty* in (2b) is identical to that of *slightly dirty* (including instances with slight amount of dirt), so B's utterance is understood as confirming A's. By contrast, the coarse interpretation of *dirty* in (2a) excludes slightly dirty entities (slight dirt is ignorable), so in (2a), B and A's utterances are conceived as a disagreement. Thus, we predict lower agreement rates in coarse-to-fine contexts like (2a) than fine-to-coarse ones, opposite to the prediction for numerals (fine-coarse >> coarse-fine).

2) a. A: The table is dirty. B: *No (?Yes), it's slightly dirty.*
   b. A: The table is slightly dirty. B: *Yes (?No), it's dirty*.

Experimental evidence. 156 participants were recruited via Amazon Mechanical Turk, with 25 different participants per question. The stimuli consisted of 60 texts with numerals, divided to 15 4-item blocks, with 3 types of round numbers (10, 100 or 1000), and 5 types of precise numbers (e.g., 99.23, 99.2, 99, 98.8, 98.77 for 100), differing in precision (coarse, fine, and very fine, depending on the number of non-zero digits), and distance from round number. Examples (3) illustrate the 5 types of precise numbers for the round number 100 in a round-to-precise sequence. Each text occurred in two versions – two inference patterns: "If round, precise" and "If precise, round").

3) a. Middle distance, coarse precision ("If 100, 99"): Nick thinks there are 100 balloons in the sky. Nick's mother thinks there are 99 balloons in the sky. Would Nick agree that there are 99 balloons?
   b. Short distance, fine precision ("If 100, 99.2"): Nick thinks the table's width is 100 cm. Nick's mother thinks it is 99.2 cm. Would Nick agree that it is 99.2 cm?
   c. Long distance, fine precision ("If 100, 98.8"): Nick thinks Sara jumped a distance of 100 cm. Nick's mother thinks it was 98.8 cm. Would Nick agree that it was 98.8 cm?
   d. Short distance, very fine precision ("If 100, 99.23"): Nick thinks George's height is 100 cm. Nick's mother thinks it is 99.23 cm. Would Nick agree that it is 99.23 cm?
   e. Long distance, very fine precision ("If 100, 98.77"): Nick thinks Bill's weight is 100 kilos. Nick's mother thinks it is 98.77 kilos. Would Nick agree that it is 98.77 kilos?

Participants were asked to provide an answer on a scale ranging from 1 (certainly not) to 7 (certainly yes). Additionally, 120 fillers involved inference relations "If A, M A" and "If M A, A" with total, partial, or relative adjectives A, and *slightly* or *completely* as modifiers M. The following examples (a,b) illustrate the fillers.

4) a. Nick thinks that Ann, his sister, is sick. Nick's mother thinks that she is completely sick. Would Nick agree that Ann is completely sick?
   b. Nick thinks that the road is slightly bumpy. Nick's mother thinks that it's bumpy. Would Nick agree that it's bumpy?

The results for numerals confirmed the prediction: coarse-fine >> fine-coarse (Wilcoxon z=-6.47, p<.0001), suggesting that, e.g., if "about 100", speakers tend to agree that "99", but if "exactly 99", they rather disagree that "exactly 100". A 2-way factorial anova for 2 randomized blocks (inference type: coarse-fine vs. fine-coarse) and a repeated measure (round number type), yields a significant inference type effect (df=1, F=83.9, p<.0001), and round number effect (df=2, F=40.8, p<.0001), though no significant interaction (df=2, F=0.54, p>.05). A 2-way factorial anova for 2 blocks (inference type) and a repeated measure (precise number type) yields also a precise number effect (df=4, F=2.9, p<.05), with no significant interaction with inference type (df=4, F=0.2, p>.05). Thus, the inference effect is identical for the 3 tested round numbers and the 5 respective precise numbers per round number.

The results for adjectives confirmed the reversed prediction (fine-coarse >> coarse-fine) revealing significant effects of inference ("If A, M A" << "If M A, A", p<.0001), modifier type (*slightly* << *completely*, p<.0001), and adjective type (total << partial, p<.0003), and an interaction: the inference effect is stronger with *completely* than *slightly*, and is, predictably, not significant in *slightly* + total adjectives (because the precise interpretation of, e.g., *full* is different from that of *slightly full*). The results suggests that speakers tend to agree that, e.g., if "slightly bumpy", "bumpy" to a larger extent than they tend to agree that if "bumpy", "slightly bumpy", and an even stronger effect, speakers tend to agree that, e.g., if "completely flat", "flat" to a larger extent than they tend to agree that if "flat", "completely flat".

To conclude, these results support the hypotheses derived based on Lewis's (1979) constraints on granularity shifting, Krifka's (2007) representation of granularity in numerals, and its extension to adjectives. Thus, they support the view that general principles guide the interpretation of granularity and approximation in different types of scalar expressions. A crucial fact underlying the different inference effect in adjectives and numerals is that on fine interpretations different numerals denote different points, while modified and unmodified adjective forms often denote highly similar intervals. Relatedly, upper-bounded readings are dominant in numerals but appear to be minor in adjectives (see also Doran et al 2009). Future work should address the connections between level of fit and scalar implicatures and their respective role in explaining inference data.