nature human behaviour

ARTICLES

https://doi.org/10.1038/s41562-020-0924-8

Check for updates

Cultural influences on word meanings revealed through large-scale semantic alignment

Bill Thompson [1]✉, **Seán G. Roberts** [2,3] **and Gary Lupyan** [4]

If the structure of language vocabularies mirrors the structure of natural divisions that are universally perceived, then the meanings of words in different languages should closely align. By contrast, if shared word meanings are a product of shared culture, history and geography, they may differ between languages in substantial but predictable ways. Here, we analysed the semantic neighbourhoods of 1,010 meanings in 41 languages. The most-aligned words were from semantic domains with high internal structure (number, quantity and kinship). Words denoting natural kinds, common actions and artefacts aligned much less well. Languages that are more geographically proximate, more historically related and/or spoken by more-similar cultures had more aligned word meanings. These results provide evidence that the meanings of common words vary in ways that reflect the culture, history and geography of their users.

Everyday words, such as 'red', 'sad', 'house', 'run' and 'sister', may strike us as denoting concepts that exist independently of any language. In a traditional view, words such as these map onto conceptual categories that we acquire independently of experience with any language[1–4]. In a strong version of this universalist view, word meanings are "more-or-less straightforward mappings from a pre-existing conceptual space, programmed into our biological nature: humans invent words that label their concepts."[5]. Alternatively, vocabularies of different languages may reflect different solutions to categorizing objects, relations, actions and abstract ideas[6–9]. In this alternative, relative, perspective, language vocabularies are culturally evolved sets of categories that we learn during the course of learning a language[10]. Rather than reflecting an innate store of concepts, or simply mapping onto categories extracted by a common perceptual system—"The categories and types that we isolate from the world … we do not find there because they stare every observer in the face, [but because they are organized by] linguistic systems…"[11].

The universalist and relative perspectives make some of the same predictions: they both predict that languages may have many 'culture-bound'[12] words that have no translation equivalents in another language. It is unsurprising that regional animals and natural features (such as 'kangaroo' and 'fjord'), specialized artefacts ('carburetor'), technical terms ('methylation') and complex social constructs ('sabbath') may be absent from certain languages. It is precisely because of the non-universality of such meanings that languages tend to borrow words that denote them wholesale from other languages[13,14]. Both perspectives similarly enable vocabularies to adapt to differences in communicative need[15]. To the extent that people in colder climates are more likely to need to distinguish between 'ice' and 'snow', we should find that languages spoken in colder climates are more likely to lexicalize this difference and, indeed, we do[16] (see also refs. [4,17–19]).

However, when it comes to common everyday meanings, the predictions of the universalist and relative perspectives diverge. The universalist perspective predicts that words that denote common animals and artefacts, common natural features (for example, 'river' and 'sand'), basic emotions, body parts and common actions—meanings that should be similarly available to everyone regardless of the language they speak—should closely align across languages. On the whole, concrete terms may be expected to vary less (that is, align better) than abstract terms. Differences, where they are found, should be random and unpredictable. By contrast, the relative view (in agreement with the intuitions of many lexicographers[12,20]) predicts that words that denote even highly concrete and seemingly self-evident meanings may fail to align across languages. Importantly, the degree of alignment should be predictable from cultural, historical and geographical factors. Languages that are geographically closer, have more recent common ancestors and are spoken by more culturally similar groups should have words that are more alignable in meaning.

Here we examined which semantic domains (for example, animals, emotions, body parts and numbers) show the most and least alignment between different languages, and whether alignment is greater for more concrete terms, as predicted by the universalist view. We then examined how alignment varies for different parts of speech and how alignment relates to lexical factors, such as frequency and neighbourhood density. Finally, we examined whether the alignment between one language and another is related to cultural and historical relatedness of the two languages.

What does it mean for two words to mean 'the same thing'? Semantic equivalence can be defined in functional terms: the meaning of a word w_1 in one language (L_1) is aligned with a word w_2 in another language (L_2) if the two words are used in the same contexts by L_1 and L_2 speakers. One reason why describing the semantic structure of natural languages is difficult is that word meanings, similar to other psychological constructs, are not directly observable[21]. The most direct way to assess semantic equivalence would be to query L_1–L_2 speakers of multiple languages about the meanings of different words[22–26]. For example, the English word 'impressed' has an unambiguously positive valence[27], whereas the valence of its Italian translation equivalent, 'impressionato', is relatively more negative[28]. This difference in valence suggests that 'impressed' and 'impressionato' do not quite mean the same thing. However, this approach

[1]Department of Computer Science, Princeton University, Princeton, NJ, USA. [2]School of English, Communication and Philosophy, Cardiff University, Cardiff, UK. [3]Department of Anthropology and Archaeology, University of Bristol, Bristol, UK. [4]Department of Psychology, University of Wisconsin-Madison, Madison, WI, USA. ✉e-mail: wdt@princeton.edu

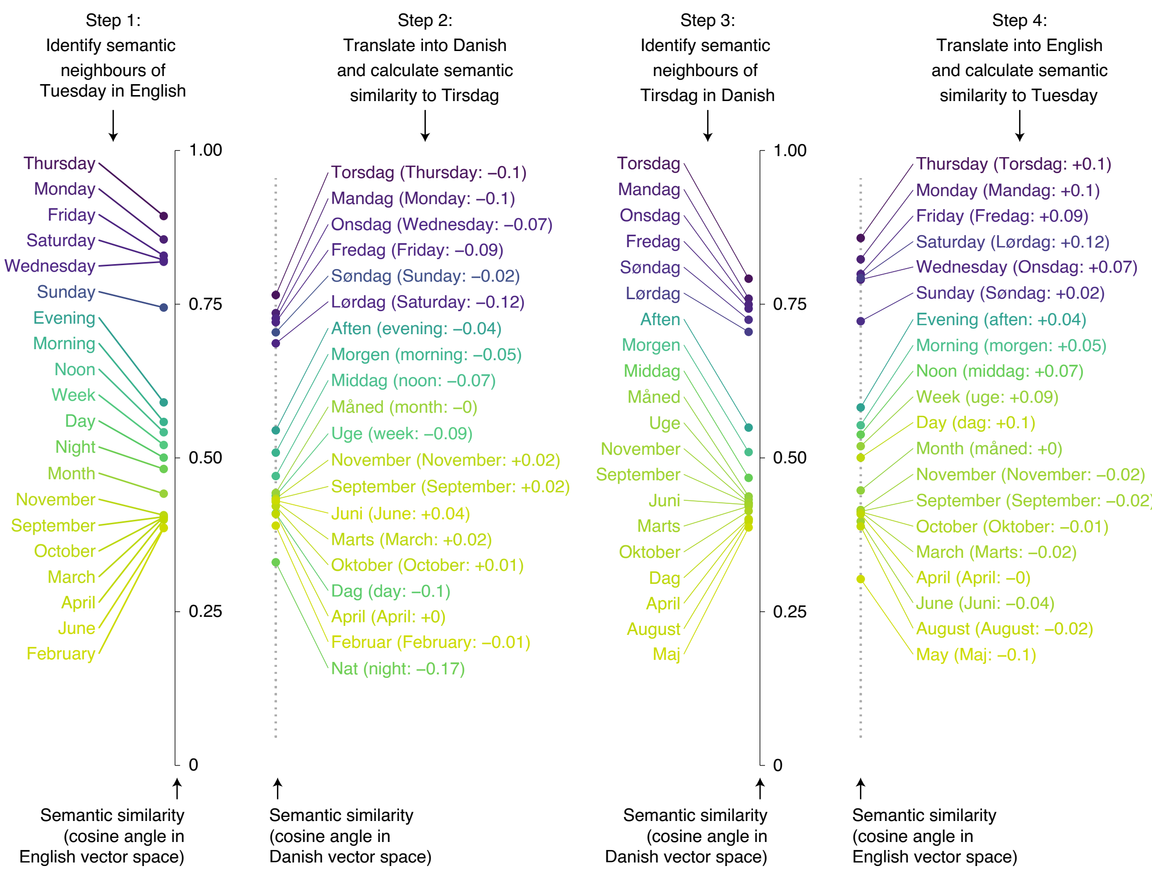

Fig. 1 | High alignment between English ('Tuesday') and Danish ('Tirsdag'). A schematic of the algorithm for computing semantic alignment. The colour denotes semantic similarity in the first language; similar colour ordering on both sides of the plot indicates a high level of alignment.

is difficult to implement at scale. For this reason, existing attempts to quantify semantic structure have focused on comprehensively analysing a specific language (often English[29,30]), cross-linguistic comparisons of a small set of meanings[31–35] or a single domain such as emotion words[36].

Here we present a large-scale analysis of vocabularies spanning 1,010 distinct 'concepts' in 41 languages (a list of the concepts is provided in Supplementary Table 1.4.2.3). Our analysis specifically focuses on words that are, in principle, highly translatable, and takes advantage of recent advances in distributional semantics. Distributional semantics is based on the idea that it is possible to understand the meaning of words by observing the contexts in which they are used—"you shall know a word by the company it keeps"[37]. The meanings of w_1 and w_2 are similar to the extent that the contexts in which they are used are similar. Early attempts to quantify semantic similarity using only contextual information were surprisingly successful at learning reasonable semantic embeddings[38–40] but were hampered by computational intractability. Advances in machine learning[41], combined with the availability of large corpora of digitized text, have now made it possible to estimate representations of word meanings—word embeddings—in a manner that correlates with human semantic judgments with a surprising degree of subtlety[42–51].

Semantic representations derived from word embeddings capture both the range of contexts in which a word is used and the relative frequencies of those contexts. Comparing contexts of use across languages enabled us to quantify, in a data-driven manner, what is sometimes called 'partial equivalence'[52]—similarities and differences in the semantic profiles of translation-equivalent pairs. If word meanings reflect self-evident natural partitions or universal constraints on how people form concepts, we should find substantial regularities in these semantic profiles across languages. For example, if the English meanings of 'in' and 'out' depend on categories that are embedded in the physical world, for example, inward motion/outward motion as perceived by all humans, then translation equivalents of the English 'in' and 'out' are likely to be used in all (or most) of the same contexts, yielding high alignment between these terms and their translation equivalents in other languages.

We obtained word forms for 1,010 concepts in 41 languages using the NorthEuraLex (NEL) dataset[53]. NEL is compiled from dictionaries and other linguistic resources that are available for individual languages in Northern Eurasia. Translation pairs can be derived from NEL because it provides word forms for the same set of concepts in multiple languages. For example, NEL provides word forms for the concept 'DOG' in 107 languages (for example, English, 'dog'; French, 'chien'; and Finnish, 'koira'). Each of the NEL concepts can be assigned to a semantic domain (for example, the concept DOG is assigned to the semantic domain 'Animals', whereas the concept NOSE was assigned to 'the body') using the Concepticon organizing scheme (see Methods).

In our main analyses, we analysed word embeddings derived from applying the fastText skipgram algorithm to language-specific versions of Wikipedia[54]. We also replicated these analyses using embeddings derived from the OpenSubtitles2018 database[55] and from a combination of Wikipedia and the Common Crawl dataset[56]. Details of these replications (Supplementary Information, section 1.2.2) and others, including an analysis of alignment using a much larger set of translation equivalents (Supplementary Information, section 1.2.1) and an analysis of how our alignment measure relates to alignment derived from patterns of colexification (Supplementary

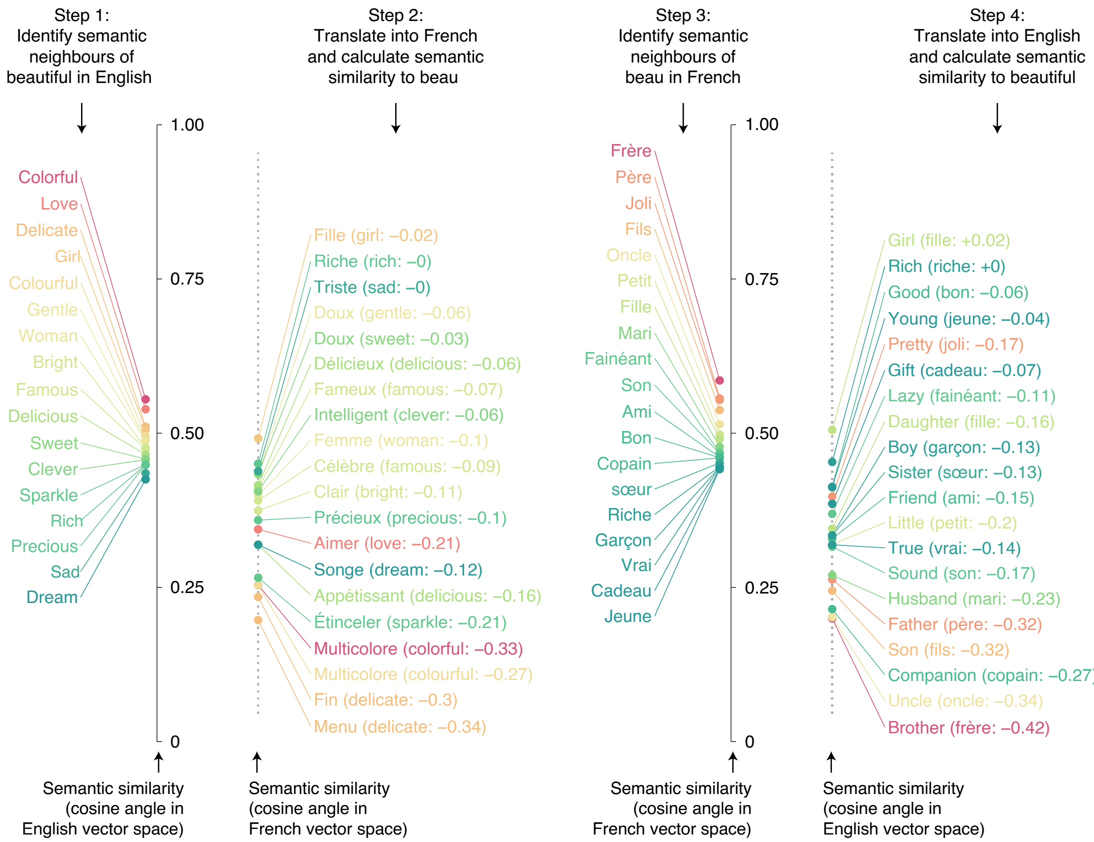

Fig. 2 | Low alignment between English ('beautiful') and French ('beau'). A schematic of the algorithm for computing semantic alignment. The colour denotes semantic similarity in the first language; perturbed colour ordering indicates a low level of alignment.

Information, section 5.3.3), are provided in the Methods and the Supplementary Information.

Figures 1 and 2 schematize our alignment algorithm. For a given language pair (L_i and L_j) and concept (c), we first identified the closest k semantic neighbours of the word for c in the vector embeddings of L_i (restricted to words that can be translated into L_j; in our primary analyses, this means that semantic neighbours are limited to the NEL vocabulary; analyses of larger translation vocabularies are provided in section 1.2.1 of the Supplementary Information, and details of how our method deals with NEL concepts that are associated with multiple words are provided in section 1.1 of the Supplementary Information). We then determined whether the translations of these neighbours are also close semantic associates of the word for c in L_j. The directional semantic alignment $L_i \rightarrow L_j$ is the Pearson correlation between these sets of similarities in both languages. For example, in Fig. 2, the closest neighbours to the American English word 'beautiful' are 'colorful' (0.55), 'love' (0.53) and 'delicate' (0.51). French translations of these neighbours are more distant from 'beau', ('multicolore' = 0.22, 'aimer' = 0.32 and 'fin' = 0.2). This reduces the correlation, so alignment is low in this direction (alignment is lowest when neighbour similarities are uncorrelated). The procedure was then repeated in the opposite direction—the k closest semantic neighbours to the word for c in L_j were identified and matched to their translations into L_i; the same Pearson correlation statistic was calculated for $L_j \rightarrow L_i$. The semantic alignment of c is the average of the two correlations. We refer to this quantity as a. Alternative measures of alignment are discussed in section 1.4 of the Supplementary Information.

We used this algorithm to analyse semantic alignment in a dataset (see Methods) that includes 1,010 concepts across 21 semantic domains (for example, kinship, animals and body parts) with an average of 48 concepts per domain (median = 40, minimum = 12, maximum = 136) in 781 language pairings from 41 languages, spanning 10 language families, with an average of 797 concepts per language pair (median = 837, minimum = 67, maximum = 991).

Results

Validating computed semantic alignment. Does lower semantic alignment correspond to words that mean different things in different languages? We validated that our alignment measure tracks differences in translatability using several methods. First, we obtained human-rated translation similarity for 201 Dutch–English translation pairs in our dataset[24]. Computed alignment was significantly correlated with Dutch–English translation similarity judgments ($r = 0.33$, $P < 0.001$). This moderate correlation increased to $r = 0.60$ ($P = 0.02$) when aggregated by the 15 semantic domains that contained ratings for at least five words, and remained a significant predictor when controlling for semantic domain ($b = 0.14$, 95% confidence interval (CI) = 0.078–0.203, $t = 4.37$, $P < 0.001$) and differences in log-transformed word frequency ($b = 0.13$, 95% CI = 0.068–0.194, $t = 4.08$, $P < 0.001$). We further confirmed the positive relationship between computed semantic alignment and human ratings using a set of Japanese–English translatability ratings[26] for 192 word pairs. These were also significantly correlated with alignment ($r = 0.29$, $P < 0.001$); furthermore, we achieved a nearly identical result using an independent set of machine-generated translations to compute alignment ($r = 0.30$, $P < 0.001$).

As an additional validation, we used our semantic alignment measure to predict differences in name agreement for 750 images each named by speakers of six languages (Spanish, English, French, Italian,

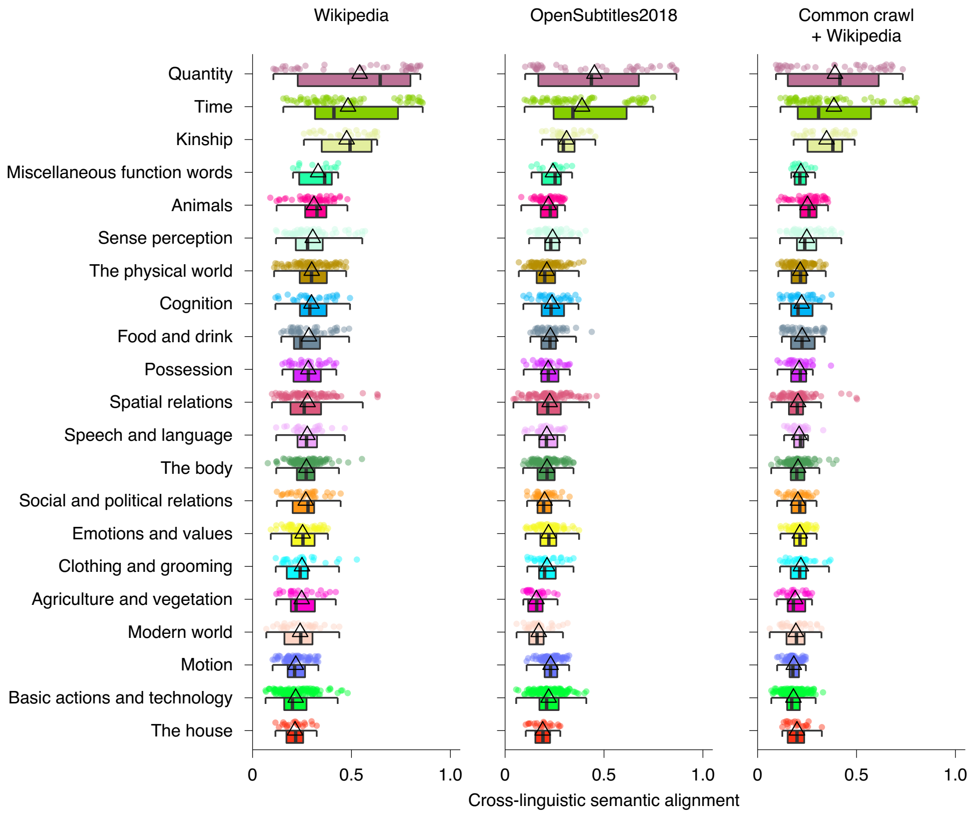

Fig. 3 | Semantic alignment of 21 semantic domains. Semantic domains are ranked according to mean cross-linguistic semantic alignment, as computed from word embeddings induced from Wikipedia (left), OpenSubtitles 2018 (middle), and Wikipedia and Common Crawl (right). Each point shows the semantic alignment for a unique pair of languages averaged over all translation pairs in the relevant semantic domain. For the box plots, the centre line shows the median, the box limits show the upper and lower quartiles, and the whiskers show 1.5× the interquartile range. The triangles show mean alignment.

German and Netherlands Dutch)[57]. Unsurprisingly, some images are named more consistently (for example, cat and gloves) than others (such as megaphone and clothes drying rack). We expected that meanings with lower semantic alignment will correspond to less consistent patterns of name agreement across the six languages. Overall, images with lower name agreement (for example, '(clothes) hanger' and 'gym') corresponded to words with lower overall alignment between these six languages (although the correlation is relatively small; $r=0.17$, $P<0.001$). Interestingly, whereas some images had high name agreement in all six languages, other images had high agreement in some languages but not in others. For example, an image of a clothes hanger has high agreement in Spanish (100% produce 'percha'), less in English (77% produce 'hanger') and even less in Italian (only 33% produce the modal response, 'appendino'). We predicted that such differences in agreement would be associated with lower alignment. Confirming this prediction, larger differences in name agreement were associated with lower alignment ($b=-0.20$, 95% CI $=-0.256$ to -0.146, $t=-7.21$, $P<0.001$). This relationship continued to remain reliable when adjusting for cross-linguistic differences in log-transformed word frequencies, as well as when taking into account the geographical and historical relationships between languages ($b=-0.13$, 95% CI $=-0.190$ to -0.071, $t=-4.28$, $P<0.001$).

Comparing alignment in 21 semantic domains. As shown in Fig. 3, alignment varied by semantic domain. On the universal perspective, alignment is predicted to be greatest for words denoting natural kinds and highly concrete meanings, such as common artefacts. Our analysis did not reveal support for this prediction. There was no statistically significant relationship between concreteness (derived from English-based norms[58]) and alignment ($t=-0.980$, $P=0.33$). Some natural kind terms were relatively well aligned, for example, 'dog' ($a=0.37$), 'wind' ($a=0.38$) and 'water' ($a=0.28$). As a benchmark, we calculated the alignment of NEL concepts in English from two different corpora and found that the average alignment was $a=0.53$ (maximum $=0.98$ for 'thirty', minimum $=0$ for 'rustle', Supplementary Information, section 1.3.2; further baseline analyses of cross-linguistic alignment are provided in section 1.3.1 of the Supplementary Information). In light of this within-language expectation, terms such as 'dog' and 'food' were well aligned across languages. However, other natural kind terms had among the lowest alignments—for example, 'feather' ($a=0.12$) and 'branch' ($a=0.12$). Similarly, words pertaining to universal aspects of human existence showed variability in alignment, such as 'move' ($a=0.14$), 'sad' ($a=0.32$) and 'food' ($a=0.42$). The most-aligned words were instead number words, temporal terms ('day', 'week' and 'spring') and common kinship terms ('daughter', 'son' and 'aunt') with alignments ranging from $a=0.49$ to $a=0.84$. Alignments for all 1,010 meanings are reported in Supplementary Table 1.4.2.3.

The meanings that are most alignable (for example, numbers and kinship terms) stand out not as being especially concrete or reflecting 'natural' joints, but as domains that have high internal coherence. Although kinship systems vary, terms denoting close kin relations are organized along a few dimensions, such as gender (son/daughter, mother/father) and generation (grandmother/mother/daughter)[59,60]. This low dimensionality seems to enable high alignment. Similarly, although a base-ten counting system is a cultural invention, once adopted, it imposes strong constraints such that the semantic difference between the English words 'five'

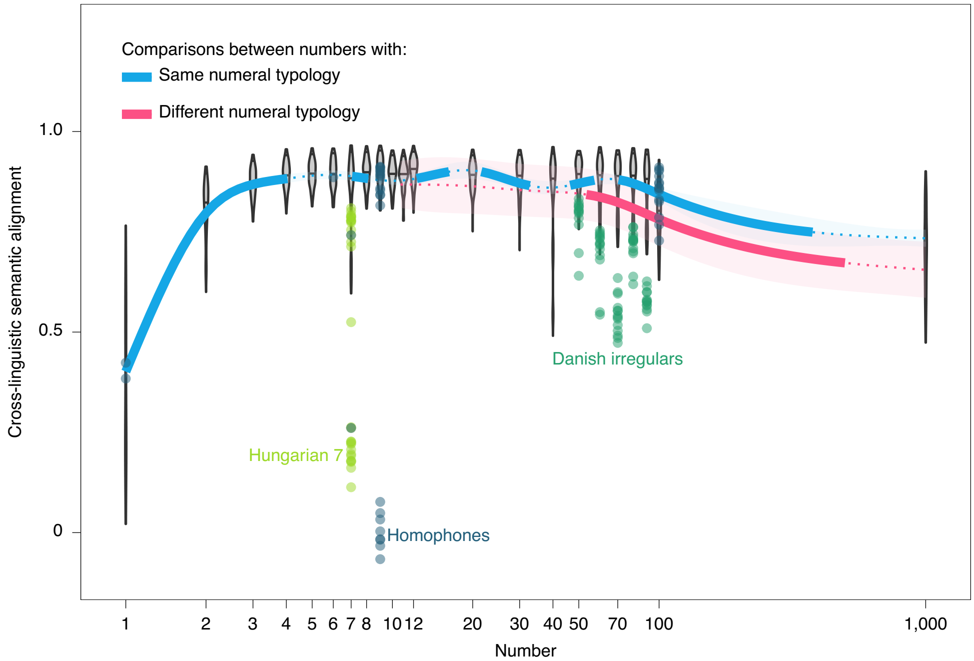

Fig. 4 | Semantic alignment of number words. Semantic alignment between 16 languages for 22 number words. Each violin plot shows the distribution of alignment values. Generalized additive model (GAM) lines are plotted for comparisons between numbers with the same numeral typology (blue) and different numeral typology (pink). The ribbons show the 95% CI around the mean and the solid lines indicate areas of significant increase or decrease. Thus, the main difference in numeral typology applies to numbers above 40. Various outliers belong to three groups: comparisons with Hungarian 7, words with alternative meanings (for example, French 'neuf' meaning '9' or 'new') and Danish irregular numbers.

and 'ten' is nearly identical to the difference between the Spanish equivalents 'cinco' and 'diez'. However, there is also systematic variation in number terms (Fig. 4 and Supplementary Information, section 4.2). The words for '1' and '2' have lower alignment than other numerals, probably because they are grammaticalized as indefinite and dual markers, respectively[61]. In general, alignment is lower for words with more polysemous meanings (for example, Hungarian 'hét' means '7' and 'week'); an analysis of how semantic alignment relates to polysemy, as quantified using colexification networks[36,62], is provided in section 5.3.2 of the Supplementary Information. Alignment is also lower for larger numbers ($P<0.001$), possibly due to their lower absolute frequency in language[63]. For numbers of 50 or higher, alignment is lower if the numbers are constructed with different numeral typologies (for example, '80' in English is constructed as 8×10 but, in French, 'quatre-vingts' is 4×20 (ref. [32]); interaction effect, $P<0.001$). These results are robust to controls for historical relatedness. Although our alignment measure is sensitive to only word co-occurrences, we can detect in the alignment patterns certain historical vestiges. For example, modern Danish uses a standard base-ten system but some number terms still reflect their historical roots in a base-twenty system (for example, 60 'tres' is 3×20) and an archaic form of 'half' (for example, 70 'halvfjerds' is 3.5×20)[32]. These irregular Danish number terms align significantly less well with the corresponding numerical terms in other languages compared with other Danish number terms ($b=-0.02$, 95% CI $=-0.025$ to -0.018, $t=-11.25$, $P<0.001$; Fig. 4).

Predicting alignment from syntactic and lexical factors. We next examined how alignment relates to several other lexical properties. 'Part of speech' was a highly significant predictor of alignment, accounting for 16% of the variance ($P<0.001$). Verbs, conjunctions and prepositions were the least aligned; 'wh' words and numerals were the most aligned (Fig. 5). There was no statistically significant interaction between part of speech and concreteness ($\chi^2_8=1.58$, $P=0.127$). Other significant predictors of alignment were absolute differences in word frequency and semantic neighbourhood density[64] (a simple measure of the extent to which words are embedded in a system of semantically related terms). Larger differences in log-transformed word frequencies[65] were correlated with lower alignment ($r=-0.20$; $b=-0.04$, 95% CI $=-0.04$ to -0.037, $t=-77.25$, $P<0.001$). Similarly, greater differences in the log-transformed semantic neighbourhood density (computed for all languages; Supplementary Information, section 1.1.5) were negatively correlated with alignment ($r=-0.13$; $b=-0.01$, 95% CI $=-0.0104$ to -0.0099, $t=-67.72$, $P<0.001$). Semantic domain, part of speech, frequency and neighbourhood density differences accounted for 30% of the variance in alignment in a mixed-effects model with language pair and concept as random effects.

Furthermore, our semantic alignment measure was strongly related to the rate at which word-forms change over time. How quickly a word form changes is not only related to its frequency[66], but also to its alignment. More aligned meanings tended to have word forms that show slower rates of change (Supplementary Information, section 1.3.4).

Predicting semantic alignment from culture and history. The relative perspective predicts that languages spoken by people with more-similar cultures should align to a greater extent. Confirming this prediction, we found that cultural similarity (the proportion of cultural traits in common on the basis of 92 non-linguistic cultural traits for 39 societies representing 39 languages in our sample[67]) predicted semantic alignment between languages ($b=0.20$, $t=6.01$, $P<0.001$). Word meanings of more-similar cultures aligned better. The same pattern was found for geographical distance ($b=-0.20$, $t=-6.42$, $P<0.001$), and for a patristic-distance-based measure (Supplementary Information, section 4.1) of language history (available only for Indo-European languages, $b=-0.178$, $t=-3.03$, $P=0.002$). Cultural similarity continued to correlate with semantic

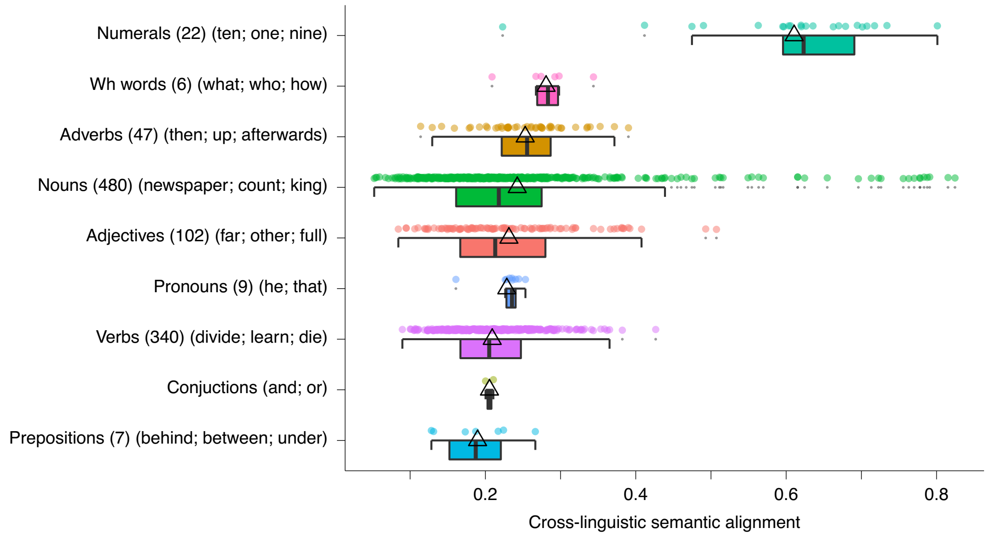

Fig. 5 | Semantic alignment by part of speech. Numerals were most strongly aligned across languages, followed by 'wh' words and adverbs. Prepositions were the least aligned. Each point is the average alignment of one concept across all pairs of languages. Words in parentheses are examples of the words included in each category. The triangles show mean cross-linguistic alignment. For the box plots, the centre line shows the median, the box limits show the upper and lower quartiles, and the whiskers show 1.5× the interquartile range.

alignment when controlling for language history and geographical proximity ($b = 0.25$, $t = 3.16$, $P = 0.002$). In these tests, we used language families and geographical area as random effects to control for non-independence of languages. Additional tests that further assess the robustness of these relationships to non-independence are provided in section 4.1 of the Supplementary Information.

Our finding that semantic alignment is predictable to a certain extent from culture, language history and geography ($R^2 = 0.363$) contrasts with previous research based on patterns of polysemy, which failed to find these relationships and has been interpreted as support for the universalist perspective[34]. We calculated a polysemy-based alignment measure (Supplementary Information, section 5.2) using a recent, large-scale database[62] of common colexifications (an approach that has been successfully used to quantify semantic alignment in the specific semantic domain of emotion vocabulary[36]). First, we established that the relationship between semantic alignment and cultural similarity is robust to controls for polysemy (Supplementary Information, section 5.3.3). Second, we examined whether colexification is related to cultural similarity (and geographical proximity) in the same manner that our distributional measure of alignment is. This polysemy-based measure of semantic alignment was not statistically significantly related to cultural similarity, and was a much weaker predictor of geographical distance compared with the distributional measure of semantic alignment (Supplementary Information, section 4.1; see Discussion).

Using our distributional approach to alignment, we also investigated the relationship between overall cultural similarity and semantic alignment within each semantic domain (Supplementary Information, section 4.1). The strongest correlations were for words related to 'food and drink' ($r = 0.29$), 'time' ($r = 0.27$), 'animals' ($r = 0.26$) and 'the body' ($r = 0.23$; adjusted P values < 0.001). The weakest correlations were for words related to 'motion', 'basic actions', 'emotion' and 'cognition' (adjusted P values > 0.1). We can also compute cultural similarity for specific cultural domains (for example, 'subsistence type', 'rituals' and 'marriage and kinship'[67] rather than semantic domains for words. Cultural similarity related to 'subsistence type' was correlated with semantic alignment in domains including 'food and drink' ($r = 0.30$), 'animals' ($r = 0.29$), 'agriculture and vegetation' ($r = 0.25$), 'clothing and grooming' ($r = 0.25$), 'social and political relations' ($r = 0.15$) and 'spatial relations' ($r = 0.10$, all adjusted P values < 0.05). These reflect well-known relationships between subsistence types and culture[5,18,68–74]. Cultural similarity related to settlement (group size, community organization and so on) was correlated with semantic alignment in domains including 'kinship' ($r = 0.28$), 'the physical world' ($r = 0.13$) and 'spatial relations' ($r = 0.11$, adjusted P values < 0.05), also reflecting previous findings[75,76]. Finally, cultural similarity related to political organization was correlated with the semantic alignment of words related to the body ($r = 0.21$, adjusted $P = 0.001$), perhaps reflecting the use of metaphors of society as a body[77,78].

For 19 Indo-European languages for which fine-grained historical and geographical proximity were available[79,80], we found that semantic alignment was significantly correlated with historical proximity ($r = 0.34$, 95% CI = 0.16–0.5, one-tailed $P = 0.01$), but not geographical proximity ($r = 0.26$, 95% CI = 0.18–0.37, one-tailed $P = 0.08$). Figure 6 shows relationships between Indo-European languages inferred solely from their semantic alignments. For these languages, we estimated the relative contribution of geography, history and culture to alignment in each semantic domain. The relative effect of historical proximity did not differ much between domains. The relationship with cultural similarity was stronger than the relationship with geographical proximity for most domains (the three largest differences were for kinship, animals and the body; the three smallest differences were for motion, possession and spatial relations). These results hint at a trade-off—the stronger the relationship with geographical proximity, the weaker the relationship with cultural similarity ($r = -0.53$, $P = 0.014$). There was no statistically significant trade-off between cultural similarity and historical proximity ($r = 0.27$, $P = 0.24$).

In summary, semantic alignment was to some extent predictable from cultural similarity and historical relationships between languages. Strikingly, semantic alignment between languages is better predicted by cultural similarity than by the geographical proximity of the populations who speak them.

Discussion

We computed semantic alignment for 1,010 meanings in 41 languages using distributed semantic vectors derived from multilingual natural language corpora. Comparing the structure of the resulting semantic spaces enabled us to measure at scale whether translation

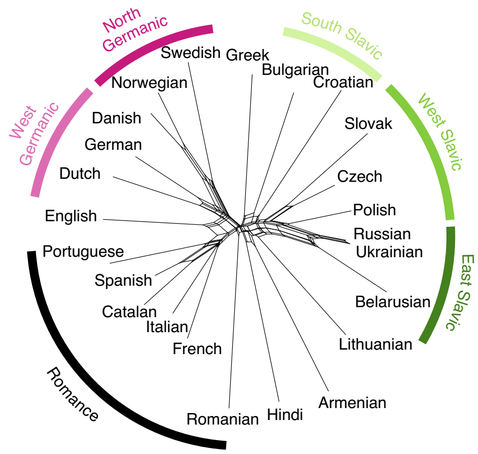

Fig. 6 | Semantic distances for Indo-European languages. Semantic distances between languages visualized as a neighbour-net (generated using Splitstree[97]; Supplementary Information, section 4.3). Distances are represented along the shortest path between language nodes. The semantic distances reflect established historical relationships, as shown by the labelling of the major sub-branches according to Glottolog[80]. Conflicting signals are shown as parallel lines. For example, English shows a conflicting signal between Germanic and Romance, which reflects its mixed history.

equivalents really mean the same thing in each language. To the extent that the vocabularies of different languages organize the world in similar ways—carving nature at its joints—their vocabularies are expected to converge on common categories and therefore should be highly alignable. By contrast, if different languages impose their own structure—carving joints into nature—word meanings may align to a more limited extent; a word and its translation equivalent may not mean quite the same thing and meanings that are easy to express in one language may not be easy to express in another.

We found that semantic alignment varies strongly with semantic domain (Fig. 3). Words for common artefacts, actions and natural kinds—meanings that should be highly aligned on a strong universalist account—were found to have only intermediate alignments. Contributing to this lower alignment were cross-linguistic differences in word frequency, differences in semantic neighbourhood densities and differences in patterns of polysemy[15,34,81].

We observed the highest semantic alignment in domains characterized by high internal structure: numbers, temporal terms and kinship terms. This result suggests that the structure of, for example, a base-ten number system and a twelve-month calendar—products of cultural evolution the acquisition of which may require experience with language[82,83]—constrains semantic relationships among words such as 'five' and 'ten', and 'month' and 'year' in a very similar manner in different languages. These high alignments may reflect a universal basis for representing these concepts[84], but the fact that alignment for kinship and temporal terms was further predicted by non-linguistic measures of cultural similarity speaks to the influence of culture on linguistic structure in these highly aligned domains. Although alignment of number words was not predicted by overall cultural similarity, alignment patterns of specific numerals for different languages were strongly influenced by the linguistic formation of these numerals (for example, whether the word for 12 is atomic (as in English) or is a composite form (for example, $10+2$, as in Bulgarian; Supplementary Information, section 4.2).

It may be tempting to ascribe our finding that words denoting common natural kinds and other everyday meanings have only intermediate alignment to random noise or other inadequacies of our method. However, the extent of such (mis)alignment between different languages was not random, but predictable from estimates of historical, geographical and cultural relatedness. Languages with greater shared history, languages that are geographically closer and languages that are spoken by more-similar cultures (as estimated from independent sources) have greater semantic alignment. This result shows that automatically derived natural language semantics contain a strong signal of cultural and historical processes.

Our reliance on corpus-derived semantic representations has limitations. Human semantic representations encode many relationships that are not present in semantics learned from word use alone[85,86]. Semantic knowledge automatically derived from corpora reflects only information contained in language and therefore under-represents what people learn about the meaning of words from direct interactions with the world. For example, the meaning of 'dog' in distributional semantic terms is derived from the contexts in which the word occurs, and includes people's direct experiences with dogs only to the extent that they are reflected in language. Although this means that distributional semantics provides an incomplete account of word meanings, distributional semantic analyses such as ours can be viewed as a conservative estimate of cross-linguistic differences in word meaning—differences that are not reflected in language use are likely to only lower estimates of semantic alignment.

Our finding that many words denoting natural kinds and common, concrete meanings show only intermediate alignment between languages, combined with the finding that this alignment is related to cultural, historical and geographical factors, conflicts with conclusions from some recent studies[34,81] in which semantic alignment was operationalized in terms of similarity of polysemy networks (that is, translation equivalents are similar to the extent that they colexify in the same manner). For example, Youn et al.[34] reported that polysemy-based alignment between geographically proximate or culturally similar languages was no greater than between randomly selected languages—"consistent with the hypothesis that cultural and environmental factors have little statistically significant effect on the semantic network of the [concepts]". This absence of an effect was interpreted as supporting the universalist view. Our view is that analyses of polysemy networks are extremely valuable in that they help us to understand how senses of words change over time. This process of change may indeed be very similar across languages[87].

However, historical regularities of sense formation do not necessarily mean that translation equivalents with similar polysemy networks mean the same thing in each language, for example, some of the senses may be much more frequent in one language or another resulting in less alignment than polysemy networks suggest. By contrast, translation equivalents with different polysemy networks do not necessarily mean very different things, for example, some of the attested senses may not be in current use. Accordingly, our analyses reveal that polysemy-based alignment, although positively correlated with our alignment measure based on distributional semantics, was only weakly related to human translatability judgments (for example, accounting for 6.5 times less variance in English–Dutch translatability human norms[24] and 2.8 times less variance in accounting for cross-linguistic differences in consistency of picture naming[57]; Supplementary Information, section 5.3). These results suggest that, of the two approaches to alignment, measures based on distributional semantics may more closely reflect differences in how words are actually used.

We were able to reproduce the absence of a relationship between polysemy-based alignment and geographical and cultural factors (Supplementary Information, section 4.1), suggesting that the difference between the current results and previous findings does not stem exclusively from differences between the sample of concepts and languages analysed, or from differences in how cultural similarity is operationalized. Our measure of alignment based on distributional models of human semantic representations was strongly associated with geographical and cultural proximity—a relationship that supports the relative position over the universalist position.

We view our research as an early attempt to quantify semantic alignment at scale using distributional semantics. Advances in machine learning—such as new methods for unsupervised alignment of vector spaces[88], and contextual word embeddings[89]—are likely to help to scale this approach even further, and address some of its limitations. Although we were able to use several datasets to validate our alignments against human data[24,26,57], and to verify that the results are robust to changes in training corpora (Supplementary Information, section 1.2.2), we recognize the need for additional validation using translatability ratings from multilingual participants. Our alignment values can be used to compile stimulus lists for these studies in ways that maximize informativeness, for example, by strategically choosing words that are predicted to have high and low translatability.

Our results do not fully fit into either the universalist or relative perspectives. The ranking of semantic domains by their alignment (Fig. 3) has unexpected elements when viewed through either the relative or the universalist lens. The finding that numerals, time, kinship and sense words have relatively high alignment may be viewed as supporting the idea that these word meanings derive from universal cognitive and perceptual biases. However, the finding that alignment is uncorrelated with concreteness and that some of the most concrete domains have relatively low alignment is unexpected on the universalist view, as are our findings that alignment of even relatively aligned domains, such as kinship and temporal terms, is strongly influenced by cultural similarity.

Our findings do not rule out the existence of universal semantic primitives into which many common words can be decomposed[90,91], although see refs. [92,93]. We think that progress in this direction probably comes from large-scale efforts focusing on aligning languages using multilingual embeddings of words and larger verbal constructions derived from naturalistic language corpora.

Methods

Data. The primary dataset that we examined (Supplementary Information, section 1.4.2) is a subset of the intersection of NEL and fast-text word embeddings trained on Wikipedia, filtered to exclude the following: any languages of which Wikipedia data do not exceed a small set of quality criteria; and any concepts that are not present in at least 20 languages. These filtering criteria (further details of which are provided in section 1.2.3 of the Supplementary Information) did not have a significant impact on our conclusions. Section 2 of the Supplementary Information provides details for all of the analyses reported in main text, including the statistical tests that we used, as well as the number of languages, language pairs, concepts and language families (these details varied between analyses because not all of the language pairs of which the alignment we calculated were available in the all of the data sources that we examined).

Notably, the 39 languages used in relating alignment to cultural similarity, geography and language history (Supplementary Information, section 4) were not a strict subset of the 41 languages used in the main analyses because restricting the sample to the languages that passed our filtering criteria reduced the overlap to 20 languages. To account for potentially low concept coverage in some of these languages, we included the number of concepts as a covariate in the models. Repeating the analyses on the 20-language subset yielded the same conclusions.

We mapped concepts listed in NEL to entries in the Intercontinental Dictionary Series (IDS) using the Concepticon[94]. The IDS is structured into chapters, which we used to assign each of the NEL concepts to a semantic domain. A full list of semantic domains and their mapping to NEL concepts is provided in section 1.4.2 of the Supplementary Information.

Algorithm. A formal description of the procedure that we used to calculate alignment for word pairs, concepts, language pairs and semantic domains is provided in section 1.1 of the Supplementary Information. Aggregate cross-linguistic alignment (at the concept and domain level) reflects simple averages taken over word-pair-level alignment in all language pairs for which relevant data were available. All analyses reported in the main text used neighbour search depth $k = 100$.

Replications. We replicated our main findings using alternative word embeddings and translation sets (including a much larger set of 20,000 translation equivalents available for a smaller number of languages[95]). We show strong correlations between our primary analyses and these replications at the level of word-pairs, concepts, language pairs and semantic domains (Supplementary Information, section 1.2). These replications justify the decision to treat alignment among the NEL concepts calculated from Wikipedia-trained embeddings as our primary dataset.

We also analysed the following: deliberately corrupted corpora to establish baseline rates of alignment (Supplementary Information, section 1.3.1); alignment between two embeddings models of a single language (English) trained on different corpora (Supplementary Information, section 1.3.2); alignment in embeddings spaces trained on lemmatized corpora (Supplementary Information, 1.2.5); alignment using different numbers of semantic neighbours (the only free parameter of our algorithm; Supplementary Information, section 1.2.2); and alignment calculated using alternative measures of structural similarity in vector spaces (Supplementary Information, section 1.1.4).

Cultural similarity. The measure of cultural similarity was based on cultural traits from the Ethnographic Atlas as linked to languages in D-PLACE[67]. Missing values were imputed by multiple imputation using classification and regression trees[96], including language family as a conditioning factor. During testing, this method imputed the correct value for held-out data 74% of the time, compared with a baseline of imputation by random choice of 19%. Cultural distances between two language groups were calculated as the average Gower distances between traits in 100 imputed sets. Further details are provided in section 4.1 of the Supplementary Information.

Historical and geographical relationships. Historical proximity was measured using patristic distances in a phylogenetic tree of 19 Indo-European languages[79]. Geographical proximity was measured as the great-circle distance between the cultural centres of each language as defined in Glottolog[80]. Further details are provided in section 4.1 of the Supplementary Information.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Data and reproducible analyses are available at https://osf.io/tngba/.

Code availability

Code to implement the alignment algorithm is available at https://osf.io/tngba/.

Received: 27 September 2019; Accepted: 2 July 2020;
Published online: 10 August 2020

References

1. Gleitman, L. & Fisher, C. In *The Cambridge Companion to Chomsky* (ed. McGilvray, J.) 123–142 (Cambridge Univ. Press, 2005).
2. Snedeker, J. & Gleitman, L. in *Weaving a Lexicon* illustrated edn (eds. Hall, D. G. & Waxman, S.) 257–294 (MIT Press, 2004).
3. Pinker, S. *The Language Instinct* (Harper Collins, 1994).
4. Berlin, B. & Kay, P. *Basic Color Terms: Their Universality and Evolution* (Univ. California Press, 1969).
5. Li, P. & Gleitman, L. Turning the tables: language and spatial reasoning. *Cognition* **83**, 265–294 (2002).
6. Evans, N. & Levinson, S. C. The myth of language universals: language diversity and its importance for cognitive science. *Behav. Brain Sci.* **32**, 429–448 (2009).
7. Lupyan, G. The centrality of language in human cognition. *Lang. Learn.* **66**, 516–553 (2016).
8. Lupyan, G. & Dale, R. Why are there different languages? The role of adaptation in linguistic diversity. *Trends Cogn. Sci.* **20**, 649–660 (2016).
9. Davidson, D. On the very idea of a conceptual scheme. *P. Am. Philos. Soc.* **47**, 5–20 (1973).
10. Lupyan, G. & Zettersten, M. in *Minnesota Symposia on Child Psychology* Vol. 40 (in the press).
11. Whorf, B. *Language, Thought, and Reality* (MIT Press, 1956).
12. Zgusta, L. *Manual of Lexicography* (Mouton, 1971).
13. Haspelmath, M. *Lexical Borrowing: Concepts and Issues. Loanwords in the World's Languages: A Comparative Handbook* 35–54 (De Gruyter Mouton, 2009).
14. Myers-Scotton, C. *Contact Linguistics: Bilingual Encounters and Grammatical Outcomes* (Oxford Univ. Press, 2002).
15. Xu, Y., Duong, K., Malt, B. C., Jiang, S. & Srinivasan, M. Conceptual relations predict colexification across languages. *Cognition* **201**, 104280 (2020).

16. Regier, T., Carstensen, A. & Kemp, C. Languages support efficient communication about the environment: words for snow revisited. *PloS ONE* **11**, e0151138 (2016).
17. Winter, B., Perlman, M. & Majid, A. Vision dominates in perceptual language: English sensory vocabulary is optimized for usage. *Cognition* **179**, 213–220 (2018).
18. Majid, A. et al. Differential coding of perception in the world's languages. *Proc. Natl Acad. Sci. USA* **115**, 11369–11376 (2018).
19. San Roque, L., Kendrick, K. H., Norcliffe, E. & Majid, A. Universal meaning extensions of perception verbs are grounded in interaction. *Cogn. Linguist.* **29**, 371–406 (2018).
20. Svensén, B. *A Handbook of Lexicography: The Theory and Practice of Dictionary-Making* 1st edn (Cambridge Univ. Press, 2009).
21. Cuyckens, H., Dirven, R. & Taylor, J. R. *Cognitive Approaches to Lexical Semantics* Vol. 23 (Walter de Gruyter, 2009).
22. Barnett, G. A. Bilingual semantic organization: a multidimensional analysis. *Journal of Cross Cult. Psychol.* **8**, 315–330 (1977).
23. Moldovan, C. D., Sánchez-Casas, R., Demestre, J. & Ferré, P. Interference effects as a function of semantic similarity in the translation recognition task in bilinguals of Catalan and Spanish. *Psicologica* **33**, 77–110 (2012).
24. Tokowicz, N., Kroll, J. F., De Groot, A. M. & Van Hell, J. G. Number-of-translation norms for Dutch–English translation pairs: a new tool for examining language production. *Behav. Res. Methods Instrum. Comput.* **34**, 435–451 (2002).
25. Dijkstra, T., Miwa, K., Brummelhuis, B., Sappelli, M. & Baayen, H. How cross-language similarity and task demands affect cognate recognition. *J. Mem. Lang.* **62**, 284–301 (2010).
26. Allen, D. & Conklin, K. Cross-linguistic similarity norms for Japanese–English translation equivalents. *Behavi. Res. Methods* **46**, 540–563 (2014).
27. Bradley, M. M. & Lang, P. J. *Affective Norms for English Words (ANEW): Instruction Manual and Affective Ratings* https://www.uvm.edu/pdodds/teaching/courses/2009-08UVM-300/docs/others/everything/bradley1999a.pdf (University of Florida, 1999).
28. Fairfield, B., Ambrosini, E., Mammarella, N. & Montefinese, M. Affective norms for italian words in older adults: age differences in ratings of valence, arousal and dominance. *PLoS ONE* **12**, e0169472 (2017).
29. Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D. & Miller, K. J. Introduction to wordnet: an on-line lexical database. *Int. J. Lexicogr.* **3**, 235–244 (1990).
30. Sigman, M. & Cecchi, G. A. Global organization of the wordnet lexicon. *Proc. Natl Acad. Sci. USA* **99**, 1742–1747 (2002).
31. Majid, A., Jordan, F. & Dunn, M. Semantic systems in closely related languages. *Lang. Sci.* **49**, 1–18 (2015).
32. Calude, A. S. & Verkerk, A. The typology and diachrony of higher numerals in Indo-European: a phylogenetic comparative study. *J. Lang. Evol.* **1**, 91–108 (2016).
33. Verkerk, A. Where do all the motion verbs come from? The speed of development of manner verbs and path verbs in Indo-European. *Diachronica* **32**, 69–104 (2015).
34. Youn, H. et al. On the universal structure of human lexical semantics. *Proc. Natl Acad. Sci. USA* **113**, 1766–1771 (2016).
35. Vivas, L., Montefinese, M., Bolognesi, M. & Vivas, J. Core features: measures and characterization for different languages. *Cogn. Process.* https://doi.org/10.1007/s10339-020-00969-5 (2020).
36. Jackson, J. C. et al. Emotion semantics show both cultural variation and universal structure. *Science* **366**, 1517–1522 (2019).
37. Firth, J. R. *Papers in Linguistics 1934-1951* (Oxford Univ. Press, 1957).
38. Lund, K. & Burgess, C. Producing high-dimensional semantic spaces from lexical co-occurrence. *Behav. Res. Methods Instrum. Comput.* **28**, 203–208 (1996).
39. Elman, J. An alternative view of the mental lexicon. *Trends Cogn. Sci.* **8**, 301–306 (2004).
40. Landauer, T. K. & Dumais, S. T. A solution to Plato's problem: the latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychol. Review* **104**, 211–240 (1997).
41. Mikolov, T., Chen, K., Corrado, G. & Dean, J. Efficient estimation of word representations in vector space. Preprint at *arXiv* https://arxiv.org/abs/1301.3781 (2013).
42. Baroni, M., Dinu, G. & Kruszewski, G. Don't count, predict! a systematic comparison of context-counting vs. context-predicting semantic vectors. in *Proc. 52nd Annual Meeting of the Association for Computational Linguistics* Vol. 1 (eds Toutanova, K. & Wu, H.) 238–247 (Association for Computational Linguistics, 2014).
43. Baroni, M. & Lenci, A. Distributional memory: a general framework for corpus-based semantics. *Comput. Linguist.* **36**, 673–721 (2010).
44. Hollis, G. & Westbury, C. The principals of meaning: extracting semantic dimensions from co-occurrence models of semantics. *Psychon. Bull. Rev.* **23**, 1744–1756 (2016).
45. Nematzadeh, A., Meylan, S. C. & Griffiths, T. L. Evaluating vector-space models of word representation, or, the unreasonable effectiveness of counting words near other words. in *Proc. 39th Annual Meeting of the Cognitive Science Society* (eds Granger, R., Hahn, U. & Sutton, R.) 859–864 (Cognitive Science Society, 2017).
46. Levy, O. & Goldberg, Y. Neural word embedding as implicit matrix factorization. *Proc. 27th International Conference on Neural Information Processing Systems* (eds Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N. D. & Weinberger, K. Q.) 2177–2185 (MIT Press, 2014).
47. Hill, F., Reichart, R. & Korhonen, A. Simlex-999: Evaluating semantic models with (genuine) similarity estimation. *Comput. Linguist.* **41**, 665–695 (2015).
48. Caliskan, A., Bryson, J. J. & Narayanan, A. Semantics derived automatically from language corpora contain human-like biases. *Science* **356**, 183–186 (2017).
49. Boleda, G. Distributional semantics and linguistic theory. *Ann. Rev. Linguist.* **6**, 213–234 (2020).
50. Tshitoyan, V. et al. Unsupervised word embeddings capture latent knowledge from materials science literature. *Nature* **571**, 95–98 (2019).
51. De Deyne, S., Perfors, A. & Navarro, D. J. Predicting human similarity judgments with distributional models: the value of word associations. in *Proc. COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers* (eds Matsumoto, Y. & Prasad, R.) 1861–1870 (COLING 2016 Organizing Committee, 2016).
52. Šipka, D. *Lexical Conflict: Theory and Practice* (Cambridge Univ. Press, 2015).
53. Dellert, J. et al. NorthEuraLex: a wide-coverage lexical database of Northern Eurasia. *Lang. Resour. Eval.* **54**, 273–301 (2020).
54. Bojanowski, P., Grave, E., Joulin, A. & Mikolov, T. Enriching word vectors with subword information. *Trans. Assoc. Comput. Linguist.* **5**, 135–146 (2017).
55. Lison, P. & Tiedemann, J. Opensubtitles2016: extracting large parallel corpora from movie and TV subtitles. In *Proc. International Conference on Language Resources and Evaluation (LREC 2016)* (eds Calzolari, N., et al) 923–929 (European Language Resources Association, 2016).
56. Grave, E., Bojanowski, P., Gupta, P., Joulin, A. & Mikolov, T. Learning word vectors for 157 languages. In *Proc. International Conference on Language Resources and Evaluation (LREC 2018)* (eds Nicoletta Calzolari, N. et al) 3483–3487 (European Language Resources Association, 2018).
57. Duñabeitia, J. A. et al. MultiPic: a standardized set of 750 drawings with norms for six European languages. *Q. J. Exp. Psychol.* **71**, 808–816 (2018).
58. Brysbaert, M., Warriner, A. B. & Kuperman, V. Concreteness ratings for 40 thousand generally known English word lemmas. *Behav. Res. Methods* **46**, 904–911 (2014).
59. Jones, D. Human kinship, from conceptual structure to grammar. *Behav. Brain Sci.* **33**, 367–381 (2010).
60. Kemp, C. & Regier, T. Kinship categories across languages reflect general communicative principles. *Science* **336**, 1049–1054 (2012).
61. Givón, T. On the development of the numeral 'one' as an indefinite marker. *Folia Linguist. Hist.* **15**, 35–54 (1981).
62. Rzymski, C. et al. The database of cross-linguistic colexifications, reproducible analysis of cross-linguistic polysemies. *Sci. Data* **7**, 13 (2020).
63. Dehaene, S. & Mehler, J. Cross-linguistic regularities in the frequency of number words. *Cognition* **43**, 1–29 (1992).
64. Vecchi, E. M., Baroni, M. & Zamparelli, R. Linear maps of the impossible: capturing semantic anomalies in distributional space. in *Proc. Workshop on Distributional Semantics and Compositionality* (eds Biemann, C. & Giesbrecht, E.) 1–9 (Association for Computational Linguistics, 2011).
65. Speer, R., Chin, J., Lin, A., Jewett, S. & Nathan, L. Luminosoinsight/wordfreq: v.2.2 (2018); https://doi.org/10.5281/zenodo.1443582
66. Pagel, M., Atkinson, Q. D. & Meade, A. Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature* **449**, 717–720 (2007).
67. Kirby, K. R. et al. D-place: a global database of cultural, linguistic and environmental diversity. *PloS ONE* **11**, e0158391 (2016).
68. Murdock, G. P. & Provost, C. Factors in the division of labor by sex: a cross-cultural analysis. *Ethnology* **12**, 203–225 (1973).
69. Sellen, D. W. & Smay, D. B. Relationships between subsistence and age at weaning in "preindustrial" societies. *Human Nat.* **12**, 47–87 (2001).
70. Apostolou, M. Bridewealth as an instrument of male parental control over mating: evidence from the standard cross-cultural sample. *J. Evol. Psychol.* **8**, 205–216 (2010).
71. Meggers, B. J. Environmental limitation on the development of culture. *Am. Anthropol.* **56**, 801–824 (1954).
72. Peoples, H. C. & Marlowe, F. W. Subsistence and the evolution of religion. *Human Nat.* **23**, 253–269 (2012).
73. Botero, C. A. et al. The ecology of religious beliefs. *Proc. Natl Acad. Sci. USA* **111**, 16784–16789 (2014).
74. Gavin, M. C. et al. The global geography of human subsistence. *R. Soc. Open Sci.* **5**, 171897 (2018).

75. Martin, M. K. & Voorhies, B. *Female of the Species* (Columbia Univ. Press, 1975).
76. Goodenough, W. H. Basic economy and community. *Behav. Sci. Notes* **4**, 291–298 (1969).
77. Lakoff, G., Espenson, J. & Schwartz, A. *The Mastermetaphor List* 2nd ed. (Univ. California Press, 1994)
78. Wiseman, R. Interpreting ancient social organization: conceptual metaphors and image schemas. *Time Mind* **8**, 159–190 (2015).
79. Bouckaert, R. et al. Mapping the origins and expansion of the Indo-European language family. *Science* **337**, 957–960 (2012).
80. Hammarström, H., Forkel, R., Haspelmath, M. & Bank, S. clld/glottolog: Glottolog database 4.2.1 https://glottolog.org/ (Max Planck Institute for the Science of Human History, 2020).
81. Srinivasan, M. & Rabagliati, H. How concepts and conventions structure the lexicon: cross-linguistic evidence from polysemy. *Lingua* **157**, 124–152 (2015).
82. Gordon, P. Numerical cognition without words: evidence from Amazonia. *Science* **306**, 496–499 (2004).
83. Tillman, K. A. & Barner, D. Learning the language of time: children's acquisition of duration words. *Cogn. Psychol.* **78**, 57–77 (2015).
84. Gelman, R. & Butterworth, B. Number and language: how are they related? *Trends Cogn. Sci.* **9**, 6–10 (2005).
85. Chen, D., Peterson, J. C. & Griffiths, T. L. Evaluating vector-space models of analogy. Preprint at *arXiv* https://arxiv.org/abs/1705.04416 (2017).
86. Huebner, P. A. & Willits, J. A. Structured semantic knowledge can emerge automatically from predicting word sequences in child-directed speech. *Front. Psychol.* **9**, 133 (2018).
87. Ramiro, C., Srinivasan, M., Malt, B. C. & Xu, Y. Algorithms in the historical emergence of word senses. *Proc. Natl Acad. Sci. USA* **115**, 2323–2328 (2018).
88. Grave, E., Joulin, A. & Berthet, Q. Unsupervised alignment of embeddings with wasserstein procrustes. Preprint at *arXiv* https://arxiv.org/abs/1805.11222 (2018).
89. Peters, M. E. et al. Deep contextualized word representations. Preprint at *arXiv* https://arxiv.org/abs/1802.05365 (2018).
90. Wierzbicka, A. *Semantics: Primes and Universals* (Oxford Univ. Press, 1996).
91. Goddard, C. & Wierzbicka, A. *Meaning and Universal Grammar: Theory and Empirical Findings* (John Benjamins Publishing, 2002).
92. Aitchison, J. *Words in the Mind: An Introduction to the Mental Lexicon* 4th edn (Wiley-Blackwell, 2012).
93. Matthewson, L. Is the meta-language really natural? *Theor. Linguist.* **29**, 263–274 (2008).
94. List, J. M., Greenhill, S., Rzymski, C., Schweikhard, N. & Forkel, R. (eds) Concepticon 2.0 (Max Planck Institute for the Science of Human History, 2019).
95. Conneau, A., Lample, G., Ranzato, M., Denoyer, L. & Jégou, H. Word translation without parallel data. Preprint at *arXiv* https://arxiv.org/abs/1710.04087 (2017).
96. van Buuren, S. & Groothuis-Oudshoorn, K. mice: multivariate imputation by chained equations in R. *J. Stat. Soft.* 1–68 (2010).
97. Huson, D. H. & Bryant, D. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23**, 254–267 (2005).

Acknowledgements

We thank J. Dellert and A. Majid. B.T. and G.L. acknowledge support from LEVINSON fellowships at the Max Planck Institute for Psycholinguistics. S.G.R. was partially supported by a Leverhulme early career fellowship (ECF-2016-435). G.L. was partially supported by NSF-PAC 1734260. The funders had no role in the conceptualization, design, data collection, analysis, decision to publish or preparation of the manuscript.

Author contributions

B.T., S.G.R. and G.L. designed the research, collected and analysed data, and contributed to the writing of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41562-020-0924-8.

Correspondence and requests for materials should be addressed to B.T.

Peer review information Primary Handling Editor: Charlotte Payne.

nature research

Corresponding author(s): Bill Thompson

Last updated by author(s): Jun 9, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection: The open-source code base for our data collection procedures is available at https://osf.io/cxqy2/

Data analysis: The open-source code base for our data analysis procedures is available at https://osf.io/cxqy2/

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data we collected and analyzed is available at https://osf.io/cxqy2/

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Quantitative data
Research sample	N/A
Sampling strategy	N/A
Data collection	N/A
Timing	N/A
Data exclusions	N/A
Non-participation	N/A
Randomization	N/A

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study	
☒	☐	Antibodies
☒	☐	Eukaryotic cell lines
☒	☐	Palaeontology and archaeology
☒	☐	Animals and other organisms
☒	☐	Human research participants
☒	☐	Clinical data
☒	☐	Dual use research of concern

Methods

n/a	Involved in the study	
☒	☐	ChIP-seq
☒	☐	Flow cytometry
☒	☐	MRI-based neuroimaging