Abstracts from the 76th Language Lunch

1st of December 2022; G.07 Informatics Forum

Exploring Contextual Code-Switching for Pretraining in Multilingual Neural Machine Translation

Vivek Iyer; Informatics; v.iyer@sms.ed.ac.uk

Multilingual pretraining approaches in Neural Machine Translation (NMT) have shown that training models to denoise synthetic code-switched data can yield impressive performance gains — owing to better multilingual semantic representations and transfer learning. However, they generated the synthetic code-switched data using non-contextual, one-to-one word translations obtained from lexicons - which can lead to significant noise in a variety of cases, including the poor handling of polysemes and multi-word expressions, violation of linguistic agreement and inability to scale to agglutinative languages. To overcome these limitations, we propose an approach called Contextual Code-Switching (CCS), where contextual, many-to-many word translations are generated using a `base' NMT model. We conduct experiments on 3 different language families - Romance, Uralic, and Indo-Aryan - and show significant improvements (by up to 5.4 spBLEU points) over the previous lexicon-based SOTA approaches. We also observe that small CCS models can perform comparably or better than massive models like mBART50 and mRASP2, based on the quantity of the data used. We empirically analyse several key factors responsible for the same - including context, many-to-many substitutions, code-switching language count etc. - and prove that these contribute to enhanced NMT pretraining on code-switched data.

Internal reconstruction of Jingpho auxiliary

Zicha Chen; PPLS; ziche.chen@outlook.com
Arturo Oncevay, Alexandra Birch; The University of Edinburgh

The Auxiliary in Jingpho (Tibeto-Burman) contains approximately 300 members. Although noticed and described by previous research, it suffers from a lack of full exploration and analysis from the perspective of phonology and morphology. This study carries out such morphophonological study and puts forward an internal reconstruction of the auxiliary system, after elucidating its inner morphological structure and functions of morphemes. The study shows that Jingpho auxiliary is in essence part of a sophisticated verbal morphology and is historically an affix complex with various morphosyntactic behaviours. These facts become most obvious after the mechanisms of phonetic reduction and analogical extension, both being neglected by the studies before, are introduced.

Uncovering Object Categories in Infant Views

Naiti S. Bhatt; PPLS; n.s.bhatt@sms.ed.ac.uk
Bria Long, Michael C. Frank; Stanford University

While adults recognize objects in a near instant, infants must learn how to categorize the objects in their visual environments. Egocentric head-mounted camera videos contain rich data that illuminate infant visual experience (Clerkin et al., 2017; Franchak et al., 2011; Yoshida & Smith, 2008).While past work has focused on the social information available to infants, here, we characterize the objects in infants' home visual environments by modifying modern computer vision models for the infant view. To do so, we collected manual annotations of objects that infants seemed to be interacting with in a set of frames from the SAYCam dataset, a longitudinal dataset of egocentric head-cam videos (Sullivan et al., 2020). We used these annotations to fine-tune region-based convolutional neural networks for object detection and segmentation (Lin et al., 2017; He et al., 2017).We found that objects in infant visual scenes lay on a right skewed Zipfian distribution, with a few objects appearing many times and most objects appearing few times. This distribution affected our model fine-tuning, attempted for 10 categories, as models trained on the skewed distribution were only able to learn a few objects well and the rest of the objects poorly. These findings and limitations help drive future work exploring infant category and language learning by elucidating the statistics of infant visual experience and tackling fine-tuning with skewed data distributions.

Development of the Shape Bias in English & Mandarin: An Online Eye-tracking Investigation of Online Processing

Manqiu Xiong; PPLS; n.s.bhatt@sms.ed.ac.uk
Ella Davis, Cara Smith, Hugh Rabagliati; University of Edinburgh, School of Psychology, Philosophy and Language Sciences

During childhood, English learners develop a shape bias for word meanings: They extend newly-learned object names based on matches in shape rather than other object features like material. Li et al. (2009) suggest that this is language-specific: They found that compared to Mandarin-speaking children, adults showed a decrease in the shape bias when extending words (Li et al., 2009). We attempted to replicate this cross-linguistic developmental pattern, using online testing and incorporating real-time processing measures (i.e., eye-tracking). Through videoconferencing, Four-year-olds and adults were shown a novel object and taught its novel name, and were then asked if the name generalized to a test object that only matched in shape or one that matched in material. We replicated the key cross-linguistic difference: With age, English speakers were more likely to select shape-match stimuli and Mandarin speakers were less likely to select shape-match stimuli. We did not find differences in the attention pattern in shape bias between English- and Mandarin-speaking children, whereas English-speaking adults attended to shape more than Mandarin-speaking adults. These results suggest that speakers of English, a count/mass language, are more likely to assume the noun refers to the object kind which share similar shape than are speakers of Mandarin, a classifier language, and this tendency increases with age. The eye-tracking evidence suggests that the attentional processing in shape bias is also language-specific, which is driven by language experience.

Sociolinguistics of Sub-Roman Britain - Social Impact on British Latin Shift

Jack Heitman; PPLS; J.T.Heitman@sms.ed.ac.uk
John Joseph, Pavel Iosad; University of Edinburgh

When and how British Latin (BL) died (while continental Romance thrived), the extent of bilingualism in the Roman/ sub-Roman period, and the fate of the Britons (the paucity of Celtic influence on Old English (OE)) is much debated. In order to answer and reconcile these unknowns, I propose to explore the possibility that BL did not initially disappear, and that the early British population who shifted to Germanic were not wholly, or perhaps even mainly, Celtic-speaking, but that, in at least the SE Lowland zone, BL was alive leaving traces of substratum influence on vernacular OE.

I must first investigate the nature of cross-linguistic impact which depends on social situations and mechanisms of language shift. I will use evidence from curse tablets and inscriptions (Tomlin, Charles-Edwards 2013, Blänsdorf 2012), as well as comparative studies in contemporary societies with L2 bilingualism, the gender paradox, social networks, and language shift within families to mirror the language environment of sub-Roman Britain. Central to findings will be presenting the ability of modern sociolinguistic theories to be applied to the pre-modern era.

This will shed light on critical implications in uncovering potential unidentified substratum BL (phonological) influence of OE, in addition to gaining a more accurate view of when, why, and how BL died as a vernacular. Expected findings include:

1. Which social networks the L2 British women participated in, and how and to whom was the Latin influence spread.
2. The tendency of women to use conservative language to gain benefit in their social network influencing the strength of substratum influence of BL on OE, and on the other hand, their tendency towards innovation allowing for rapid assimilation to Latin and OE (Trudgill 1972, Labov 2002).
3. Rapid ethnolinguistic growth and bilingualism of Latin, and rapid ethnolinguistic decline in transition to OE in the sub-Roman period (Spolsky 2012).

Development of the Shape Bias in English & Mandarin: An Online Eye-tracking Investigation of Online Processing

Manqiu Xiong; PPLS; n.s.bhatt@sms.ed.ac.uk
Ella Davis, Cara Smith, Hugh Rabagliati; University of Edinburgh

During childhood, English learners develop a shape bias for word meanings: They extend newly-learned object names based on matches in shape rather than other object features like material. Li et al. (2009) suggest that this is language-specific: They found that compared to Mandarin-speaking children, adults showed a decrease in the shape bias when extending words (Li et al., 2009). We attempted to replicate this cross-linguistic developmental pattern, using online testing and incorporating real-time processing measures (i.e., eye-tracking). Through videoconferencing, Four-year-olds and adults were shown a novel object and taught its novel name, and were then asked if the name generalized to a test object that only matched in shape or one that matched in material. We replicated the key cross-linguistic difference: With age, English speakers were more likely to select shape-match stimuli and Mandarin speakers were less likely to select shape-match stimuli. We did not find differences in the attention pattern in shape bias between English- and Mandarin-speaking children, whereas English-speaking adults attended to shape more than Mandarin-speaking adults. These results suggest that speakers of English, a count/mass language, are more likely to assume the noun refers to the object kind which share similar shape than are speakers of Mandarin, a classifier language, and this tendency increases with age. The eye-tracking evidence suggests that the attentional processing in shape bias is also language-specific, which is driven by language experience.

Recursive Neural Networks with Bottlenecks Diagnose (Non-)Compositionality

Verna Dankers; Informatics; s2112866@ed.ac.uk
Ivan Titov; University of Edinburgh

A recent line of work in NLP focuses on the (dis)ability of models to generalise compositionally for artificial languages.However, when considering natural language tasks, the data involved is not strictly, or locally, compositional. Quantifying the compositionality of data is a challenging task, which has been investigated primarily for short utterances.
We use recursive neural models (Tree-LSTMs) with bottlenecks that limit the transfer of information between nodes. We illustrate that comparing data's representations in models with and without the bottleneck can be used to produce a compositionality metric. The procedure is applied to the evaluation of arithmetic expressions using synthetic data, and sentiment classification using natural language data. We demonstrate that compression through a bottleneck impacts non-compositional examples disproportionately and then use the bottleneck compositionality metric (BCM) to distinguish compositional from non-compositional samples, yielding a compositionality ranking over a dataset.

Regularisation of linguistic variation arises from both learning and production biases

Aislinn Keogh; PPLS; a.k.keogh@sms.ed.ac.uk

Experimental and computational work (e.g. Ferdinand et al. 2019, Smith and Wonnacott 2010, Hudson Kam and Newport 2009, Reali and Griffiths 2009) has demonstrated that individuals have cognitive biases which work against unpredictable variation. But even predictable variation is sometimes lost as languages evolve, with previously irregular forms adopting the regular pattern. What, if any, cognitive biases can be 'read off' this kind of language change? Recent work has explored the role of memory retrieval during language production as a mechanism for the loss of unpredictable variation (Hudson Kam & Chang 2009). Here, we further test this hypothesis, providing experimental evidence that regularisation of both predictable and unpredictable variation arise via this same mechanism.

Participants in an artificial language learning experiment were trained on either probabilistically conditioned or unconditioned variation in nominal plural marking, with working memory taxed during learning, production, or not at all. They were then asked to produce labels for the same set of nouns, and finally, to estimate how often they had seen each noun with each plural in training. To tax working memory, we used a concurrent digit sequence recall task (Perfors 2012).

Our results show that, while participants accurately encoded their input in terms of overall token frequency distribution, those exposed to random variation believed that their input in fact followed a pattern of lexical conditioning. Furthermore, while only participants subject to the memory load manipulation during production significantly reduced overall variability (by boosting the token frequency of one variant at the expense of the other), participants across the board produced more lexical conditioning than they were exposed to.

Overall, the results of this study support the claim that cognitive biases leading to regularisation target both unpredictable and predictable variation. Our findings support the idea that regularisation is particularly strong during production, and is driven at least in part by memory limitations. However, we also find evidence that some aspects of regularisation arise from learning biases.

Analyzing Acoustic Word Embeddings from Pre-trained Self-supervised Speech Models

Ramon Sanabria; Informatics; ramon.sanabria.teixidor@gmail.com
Hao Tang, Sharon Goldwater; University of Edinburgh

Given the strong results of self-supervised models on various tasks, there have been surprisingly few studies exploring self-supervised representations for acoustic word embeddings (AWE), fixed-dimensional vectors representing variable-length spoken word segments. In this work, we study several pre-trained models and pooling methods for constructing AWEs with self-supervised representations. Owing to the contextualized nature of self-supervised representations, we hypothesize that simple pooling methods, such as averaging, might already be useful for constructing AWEs. When evaluating on a standard word discrimination task, we find that HuBERT representations with mean-pooling rival the state of the art on English AWEs. More surprisingly, despite being trained only on English, HuBERT representations evaluated on Xitsonga, Mandarin, and French consistently outperform the multilingual model XLSR-53 (as well as Wav2Vec 2.0 trained on English).

Improving compositional generalisation for semantic parsing

Matthias Lindemann; Informatics; m.m.lindemann@sms.ed.ac.uk
Ivan Titov; University of Edinburgh
Alexander Koller; Saarland University

In contrast to humans, neural sequence-to-sequence models struggle with structural generalisation. If asked to parse a natural language question into an executable command to retrieve information from a database (semantic parsing), a sequence-to-sequence model might, for instance, be able to construct a command to look for flights on a specified date, or for flights to a given destination, but it might struggle with constructing the command for finding flights that take place on a specified date AND to a specified destination if this structure is not represented in the training data.

Taking inspiration from grammar-based models that excel at this kind of generalisation, we present a flexible end-to-end differentiable neural model that composes two structural operations: a fertility step, introduced in this work, and a reordering step based on previous work (Wang et al. 2021). These two steps conceptually decompose the semantic parsing task into two subtasks: predicting the right predicate symbols and arranging them into the correct order. To ensure differentiability of the model, so it can be trained with standard gradient descent methods, we use the expected value of each step, which we compute using dynamic programming.

Our model outperforms sequence-to-sequence models by a considerable margin on challenging compositional splits of realistic semantic parsing tasks that require generalisation to longer examples.

In addition to this completed work, I am happy to talk about promising work-in-progress on a more efficient and flexible method to perform reordering.

Exploring Prosody Transfer in Speech Synthesis

Atli Sigurgeirsson; Informatics; atlisigurgeirsson@gmail.com
Simon King; University of Edinburgh

Some recent models for Text-to-Speech synthesis aim to transfer the prosody of a reference utterance to the generated target synthetic speech. This is done by using a learned embedding of the reference utterance, which is used to condition speech generation. During training, the reference utterance is identical to the target utterance. Yet, during synthesis, these models are often used to transfer prosody from a reference that differs from the text or speaker being synthesized.

To address this inconsistency, we propose to use a different, but prosodically-related, utterance during training too. We believe this should encourage the model to learn to transfer only those characteristics that the reference and target have in common. If prosody transfer methods do indeed transfer prosody they should be able to be trained in the way we propose. However, results show that a model trained under these conditions performs significantly worse than one trained using the target utterance as a reference. To explain this, we hypothesize that prosody transfer models do not learn a transferable representation of prosody, but rather a utterance-level representation which is highly dependent on both the reference speaker and reference text.

Areal Models of Bantu Spirantization Predict Unattested Phonologies

Brandon Kieffer; PPLS; s2319614@ed.ac.uk

Janson 2007 proposed that Bantu Spirantization (BS) is the result of an areal shift, spread via contact instead of phylogenetically. I argue against this by outlining the variation in how BS and the connected vowel merger could spread via contact, showing that such a model predicts unattested patterns.

BS involves the assibilation of Proto-Bantu (PB) stops in front of its highest vowels (*i, *u). The merger was of PB's two pairs of high vowels (*i and *ɪ, *u and *ʊ). Assibilation would have been predictable before the merger, and its phonologization made evident by the loss of the trigger. Together these operations shifted the contrast from vowels to consonants (Bostoen 2008).

Example 1 Spirantization Examples (Bastin et al. 2002, Kinyarwanda-English Dictionary)

Proto-Bantu	Kinyarwanda (JD61)	English
*gid	/ku-zir-a/	"To be taboo"
*gɪd	/ku-gir-a/	"To have"
*kund	/gu-pfund-a/	"To become more full"
*kʊnd	/gu-kund-a/	"To like"

Since assibilation only applied before the highest vowels, we know that assibilation happened first (Bostoen 2008). If the merger were first, then BS would occur before all four high vowels, making it predictable and eliminating the contrast.

Example 2: Bantu Spirantization Sound Changes

	*gid	*gɪd
Assibilate	**z**id	gɪd
V. Merge	zid	g**i**d
Rwanda	/zir-a/ Be taboo	/gir-a/ have

	*gid	*gɪd
V. Merge	gid	g**i**d
Assibilate	**z**id	**z**id
Rwanda	/zir-a/ Be taboo	*/zir-a/

My typology shows how these shifts could be passed, factoring in whether both are passed, in what order, and whether each is passed before the first language underwent the merger. This typology predicts languages which assibilate before all high vowels, which is unattested, if the ordering of the transmissions is not followed correctly. Additionally, languages with incorrect BS will continue to spread their mistake. Given the ease for

incorrect transmission and the lack of languages with it, I conclude that there was a significant gap between shifts, preventing the merger from spreading first.
Assibilation and the merger were therefore two chronologically separate changes. This is surprising since they have nearly identical isoglosses spanning much of Africa (Janson 2008). While independent yet similarly-shaped isoglosses are possible, they are not probable at such a large scale. Rather, it is more likely that BS's distribution is due to genealogical spreading, at least in part.

Learning Dependencies of Discrete Speech Representation with Neural Hidden Markov Models

Sung-Lin Yeh; ILCC; sunglin.yeh@ed.ac.uk

While discrete latent variable models have had great success in self-supervised learning, most models assume that frames are independent. Due to the segmental nature of phonemes in speech perception, modeling dependencies among latent variables at the frame level can potentially improve the learned representations on phonetic-related tasks. In this work, we assume Markovian dependencies among latent variables, and propose to learn speech representations with neural hidden Markov models. Our general framework allows us to compare to self-supervised models that assume independence, while keeping the number of parameters fixed. The added dependencies improve the accessibility of phonetic information, phonetic segmentation, and the cluster purity of phones, showcasing the benefit of the assumed dependencies.