Posters

1. Acquiring production skills through implicit learning
Giulia Boloventa, John Williams, Dept. of Theoretical & Applied Linguistics

When performing many daily tasks, we rely on finely tuned knowledge that we can use even though we are not able to verbalise it - when riding a bicycle, for instance. Language is another such area: we usually have strong intuitions about what is correct and what isn't in our native language, even though we may not be able to explain why. It has been argued that the development of such implicit knowledge is also necessary to achieve fluency when learning a second language (Krashen, 1981). Over the last decades, a growing body of research in psychology has shown that it is possible to directly acquire implicit knowledge of the underlying regularity in a sequence of stimuli while remaining unaware of it (Reber, 1967; Destrebecqz and Cleeremans, 2001), a process known as implicit learning. The application of implicit learning paradigms to second language acquisition research has shown that implicit learning can occur with a variety of linguistic features, such as orthographic (Pacton et al., 2011) or syntactic (Rebuschat and Williams, 2012) regularities, even leading to the acquisition of novel form-meaning connections (Leung and Williams, 2012). However, these studies normally test for the presence of implicit knowledge by means of comprehension tasks without looking at production skills. The aim of this study was to begin addressing that gap, by looking at whether oral production skills can develop as a direct result of implicit learning.

In the experiment, we used a rule derived from Czech, in which the correct choice of spatial preposition depended on the physical properties of a given place (open vs. bounded spaces), in an elicited oral imitation task. Participants heard English sentences, where prepositions were replaced by one of four pseudowords. Two prepositions appeared in system-based sentences, in which their alternation followed the Czech rule, the other two in sentences where it occurred at random; the same place names appeared in both sentence types. While hearing the sentences, participants saw pictures matching their content. After exposure, participants had to recall the sentences in random order, as cued by the pictures on screen. Participants made significantly fewer mistakes recalling the system-based sentences than the random ones, even though they had not become aware of the underlying rule. This suggests that they had gained implicit knowledge of the rule, and they were relying on it when reconstructing sentences during recall.

---

2. Acquisition of quantity, relevance and word learning inferences, and their relationship with Theory of Mind
Elspeth Wilson, Napoleon Katsos, Dept. of Theoretical & Applied Linguistics

Recent studies have shed light on factors that affect children's development of quantity, and particularly scalar, implicatures, e.g., the inference that by uttering "I fed some of the ducks" the speaker means *some but not all of the ducks* (Papafragou & Skordos, 2016). However, we do not know how this relates to the development of relevance implicatures or to word learning by exclusion, often claimed to be a pragmatic inference. In addition, a Gricean understanding of implicatures might expect Theory of Mind to play a crucial role in inferencing, but this is puzzling given children's apparently early pragmatic development (Breheny, 2006).

In this study, scalar, ad hoc quantity, and relevance implicatures, together with word learning by exclusion are tested in English-speaking 3-5-year-olds, using a story-based picture-selection task. Measures of Theory of Mind (unexpected contents and Sally-Anne tasks, Perner et al, 1987, and Baron-Cohen et al, 1985), vocabulary (British Picture Vocabulary Scale, Lloyd et al, 1987) and grammar (mini Test for Reception of Grammar, Bishop, 2005) are collected.

Results and initial analyses from on-going data collection (N=96) show that overall patterns confirm previous findings: children make word learning by exclusion inferences youngest, followed by ad hoc and relevance inferences (children are above chance from 3½ upwards), and finally scalars. Interestingly, older 3-year-olds in this study seem to be making relevance inferences that those in another study did not (Schulze et al, 2013). However, looking at 3-year-olds only to avoid ceiling affects, there are as yet no strong patterns of correlation between inference types, with or without TROG partialled out, nor between ToM and implicature performance. This suggests that different factors may affect the developmental trajectory of different inferences, and challenges us to review models of acquisition directly inspired by Gricean pragmatics, and the methodologies typically employed to investigate implicature development and Theory of Mind.

---

3. Automated speech-unit delimitation in spoken learner English
Russell Moore, Andrew Caines, Calbert Graham & Paula Buttery
ALTA Institute, Dept. of Theoretical & Applied Linguistics and Computer Laboratory

In order to apply computational linguistic analyses and pass information to downstream applications, transcriptions of speech obtained via automatic speech recognition (ASR) need to be divided into smaller meaningful units, in a task we refer to as 'speech-unit (SU) delimitation'. We closely recreate the automatic delimitation system described by Lee & Glass (2012), `Sentence detection using multiple annotations', *Proceedings of INTERSPEECH*, which combines a prosodic model, language model and speech-unit length model in log-linear fashion.

Since state-of-the-art natural language processing (NLP) tools have been developed to deal with written text and its characteristic sentence-like units, SU delimitation helps bridge the gap between ASR and NLP, by normalising spoken data into a more canonical format. Previous work has focused on native speaker recordings; we test the system of Lee & Glass on non-native speaker (or 'learner') data, achieving performance above the state-of-the-art. We also consider alternative evaluation metrics which move away from the idea of a single 'truth' in SU delimitation, making use of the probabilities emitted by parsers and perplexity scores from statistical language models.

We report the performance of our SU delimiter in various configurations on a spoken learner corpus, both with our new metrics and established BLEU-like and information retrieval scores. Our best performing configuration makes use of a highly weighted prosodic model and native speaker language model, demonstrating what we find most advantageous about this architecture: that its modular nature allows training on various sources, which is advantageous as the learner corpora we are interested in tend to be small resources. ASR output transcripts are unpunctuated, and therefore an automated SU delimiter allows those transcripts to be subdivided and passed on to downstream applications in usable ways. In our case, we require SUs which are useful for automated learner assessment and feedback in CALL systems. We consider what makes a meaningful speech-unit, and indicate what remains to be done beyond this work.

---

4. Automatic assessment and error detection of non-native English speech using phone distance features

K. Kyriakopoulos, M.J.F. Gales, K.M. Knill
ALTA Institute & Dept. of Engineering

With growing global demand for learning English as an additional language, there has been considerable interest in methods of automatic evaluation of spoken language proficiency for use in interactive electronic learning tools and for grading candidates for formal qualifications. In this work a system is described for automatically grading the fluency level of non-native English speakers and identifying individual pronunciation errors in their utterances, using only short samples of unstructured, spontaneous speech and without using pronunciation error labels in training.

The system builds on the baseline grader system developed at the ALTA Institute. Audio is passed through an automatic speech recogniser (ASR), and the recognised text and audio used to extract prosodic and other features, which are in turn used to train a Gaussian Process (GP) grader to assign scores to speakers and sections. A corpus of recorded answers to the BULATS English speaking test, only labelled with overall speaker and section-specific scores, is used for training. The grader is enhanced by extracting new phone distance features to represent relative phone pronunciation within a certain context (e.g. a specific speaker or utterance), without the use of pronunciation error labels. First, models are trained to represent the manner in which the speaker pronounced each of the 47 phones of English. Two different types of model are investigated: a three-emitting state Hidden Markov Model (HMM), which is standard in acoustic modelling, and a simple multivariate Gaussian, which can be thought of as a simplified, lower dimensionality version of the HMM. The K-L divergence between the models representing two phones is then calculated for each of the 1081 possible phone pairs. The resulting distances are added as grader input features, producing considerable performance enhancement for both speaker native language (L1) dependent and independent paradigms, though more so for the former.

The phone distance features can also be used to identify pronunciation errors. Individual GP scorers are trained to predict the test score given the K-L divergences for a phone pair. A pronunciation score for a candidate phone is derived by combining predicted scores for all pairs containing that phone. The phone scores are then used with ASR word confidences to identify individual phone level errors in the candidate's speech. This can be applied at the speaker or question response level. A method is defined to assess the precision of these error detection mechanisms, using crowd-sourced word-level human binary judgments of pronunciation quality.

---

5. Constrained multi-task learning for automated essay scoring

Ronan Cummins, Meng Zhang, Ted Briscoe
ALTA Institute & Computer Laboratory

Supervised machine learning models for automated essay scoring (AES) usually require substantial task-specific training data in order to make accurate predictions for a particular writing task. This limitation

hinders their utility, and consequently their deployment in real-world settings. In this paper, we overcome this shortcoming using a constrained multi-task pairwise-preference learning approach that enables the data from multiple tasks to be combined effectively.

We treat essay scoring as a ranking problem. The essays from different prompts are treated as essays from different tasks in the ranking. The ranking pairs are only generated among essays from the same prompt after the features of each essay is extracted [Yannakoudakis et al., 2011]. The algorithm behind the pairwise preference ranking is a type of linear perceptron, which is called time aggregated perceptron [Briscoe et al., 2010]. This perceptron contains a sampling mechanism to reduce the number of difference pairs, and it can still maintain the distribution of essay score unchanged in the training set. Frustratingly easy domain adaptation [Daume III, 2007] is used to do feature argumentation on the whole ranking pairs to capture the similarities and differences among these prompts. Separate linear regression is trained on each prompt to map the ranking scores back based on the grading scale of each prompt.

Our study shows that using frustratingly easy domain adaptation can improve the AES system performance on different prompts. Furthermore, contrary to the result identified by Phandi et al. [2015], we show that AES system with high performance can be built with little or no task-specific training data. We perform a detailed study of our approach on the Automated Student Assessment Prize public dataset 1 in scenarios where we have varying amounts of task-specific training data and in scenarios where the number of tasks increases. In most cases, bringing more data in the training procedure can boost the performance of the system on the testing set, even if these training and testing data come from different prompts.

---

6. Cross-linguistic influences on the acquisition of metaphorical expressions

Chris Mengying Xia, Dept. of Theoretical & Applied Linguistics

This study aims to explore possible cross-linguistic influences on the acquisition of conventionally used metaphorical expressions by Chinese learners of English. In this study, "metaphorical expressions" are the lexical items that are used to deliver meanings that depart from their literal, core meanings, such as the following examples:

a. He attacked a passenger with a stick. (literal)
b. He attacked my theory. (metaphorical)

In the view of lexical semantics (e.g. Sweetser 1990), the metaphorical expressions above should be regarded as polysemous because they have two different but closely related meanings. Previous literature on cross-linguistic influences (e.g. Jordens and Kellerman 1981) and bilingual lexicon (De Groot 1992) makes different predictions regarding the transferability of metaphorical meanings of a lexical item comparing with the literal meaning. In particular, it is not clear whether they are able to derive and/or acquire the metaphorical meaning in a non-guided way when they already acquired the literal meaning of the same lexical item.

Three different conditions are examined in the study: (1) metaphorical expressions shared between the L1 and the L2 of learners, (2) metaphorical expressions available in the L1 but not in the L2, and (3) metaphorical expressions available in the L2 but not in the L1. An acceptability judgement task with sentence correction components and confidence scales was used to examine whether the learners accept different types of metaphorical expressions, and how they "correct" the incorrect use of metaphorical expressions. A survey of psychotypology was also included to discover whether the perceived distance between English and Chinese by individual learners would influence learners' judgement of transferability of metaphorical expressions.

Results show that the acquisition of metaphorical expressions resides in between the acquisition of literal meanings of lexical items and that of idioms that are semantically opaque. The participants are able to discriminate expressions that are available in different languages, demonstrate different types of cross-linguistic influence, and select different strategies when correcting the given expressions. While participants' general proficiency is an important factor for cross-linguistic influence, it influences the acquisition of metaphorical expressions in an imbalanced way when learners encounter language-specific metaphorical expressions. An asymmetry between the acquisition of literal meaning and metaphorical meaning of a lexical item is also observed, which is shown by the lower acceptability of metaphorical expressions that are available in both the L1 and the L2 in comparison to the literal counterparts.

---

7. Crowdsourcing a multilingual speech corpus: recording, transcription and annotation of the CrowdED Corpus

Andrew Caines, Christian Bentz, Calbert Graham & Paula Buttery
ALTA Institute, Dept. of Theoretical & Applied Linguistics and Computer Laboratory

We announce the release of the CrowdED Corpus[1]: a pair of speech corpora collected via crowdsourcing, containing a native speaker corpus of English (CrowdED_english), and a corpus of German/English bilinguals (CrowdED_bilingual). Release 1 of the CrowdED Corpus contains 1000 recordings amounting to

33,400 tokens collected from 80 speakers and is freely available to other researchers under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 international licence (CC BY-NC-SA 4.0).

It is well-known that building speech corpora is a time-consuming and expensive process: one estimate puts the cost of transcription at €1 per word, before the cost of any extra annotation (Ballier & Martin 2013). We present a method of collecting speech corpora via crowdsourcing facilities, showing that we can reduce costs considerably by distributing the work among multiple online workers. We recruited participants via the Crowdee application for Android operating systems (www.crowdee.de). Recruits were prompted to respond to business-topic questions of the type found in language learning oral tests. We then used the CrowdFlower web application (www.crowdflower.com) to pass these recordings to crowdworkers for transcription and annotation of errors and sentence boundaries. Finally, the sentences were tagged and parsed using standard natural language processing tools.

All Crowdee recordings and annotated CrowdFlower transcripts are made freely available to other researchers (apc38.user.srcf.net/resources/#crowded). There are some issues of data quality, which may indeed be inevitable when collecting data from 'the crowd'. We view the Corpus as useful for investigations of first language transfer, learner-native speaker comparisons, and other as yet unanticipated purposes.

[1] *We gratefully acknowledge receipt of funding from Crowdee, CrowdFlower, and Cambridge English, University of Cambridge.*

---

8. Deep density networks with uncertainty for automatic language proficiency assessment

Andrey Malinin, ALTA Institute & Dept. of Engineering

This work considers the automatic assessment of spontaneous spoken English proficiency. Automatic spoken language assessment systems are becoming increasingly important to meet the demand for English second language learning. An automatic assessment system should be able to accurately assess the learners' ability level from spontaneous prompted speech. While automatic graders are potentially more consistent than humans, the validity of their predicted grades varies, as automatic assessment is a challenging task.

A number of automatic methods have been proposed, notably Gaussian Process (GP) based graders. GP graders offer state-of-the-art performance and provide a measure of uncertainty in the form of GP variance. As shown in previous work, the uncertainty measure is sufficiently accurate to decide which automatic grades should be rejected to be re-graded by humans. Unfortunately, GP graders offer very limited scalability due to their computational and memory requirements.

This work has two novel contributions. Firstly, we propose automatic graders based on Deep Neural Networks (DNNs) and Deep Density Networks (DDNs). Secondly, we propose assessment criteria based on Area-Under-the-Curve (AUC) to evaluate uncertainty-based rejection performance of automatic graders. On experiments conducted on data from the Business Language Testing Service (BULATS) DNNs outperform both GP graders and standard human graders in Pearson correlation with expert human graders. Interpolation of GP and DNN grades further boosts performance. However, DNNs lack a measure of uncertainty in their predictions. By adopting a DDN architecture and training in a multi-task fashion to minimize KL-divergence between both the low-variance data distribution and a high-variance noise distribution, we are able to match GP grader performance in assessment and outperform GPs in variance-based rejection, based on the AUC assessment criteria.

---

9. Evaluating multi-modal deep learning systems with micro-worlds

Alexander Kuhnle, Ann Copestake
Computer Laboratory

Multi-modal deep learning systems recently showed strong performance on the image captioning task (Karpathy and Li, 2015). At the same time, however, some more detailed investigations cast doubt on the quality of the results, or rather, whether the current evaluation practice is sufficient to test for true scene and language understanding (Hodosh and Hockenmaier, 2016; Nguyen et al., 2015).

With the aim of analysing basic linguistic/symbolic capabilities of multi-modal systems, we propose to move away from real-world photos with human-written captions, as well as the task setup of asking the system to generate captions. Regarding the latter, we more closely follow the setup of the recently introduced image question answering task (Antol et al., 2015) and take an image as a natural representation of the world against which statements can be evaluated.

In our experiments, a system is trained on pairs of images and statements about them, together with corresponding values indicating the appropriateness of the proposition given the image, while during test time previously unseen object/concept combinations of the same symbolic structure as the train instances are presented. To do well on this task, it is hence crucial for the evaluated system to learn to understand the underlying symbolic principle. Instead of real-world data, we use automatically generated abstract micro-worlds, similar to other work on formally testing the abilities of deep learning systems (Sorodoc et al., 2016; Joulin and Mikolov, 2015; Bowman et al., 2015; Weston et al., 2015; Vinyals et al.,

2015; Sukhbaatar et al., 2015). In doing so, we avoid the problem of visually noisy or otherwise ambiguous instances, and are able to more exhaustively cover the space of possible images and captions. Moreover, it enables us to control the data generation, and so investigate the learning process in a network. For instance, quantifier learning can be analysed by constructing instances specifically targeting interesting quantifier configurations (Pietroski et al., 2009).

Internally, the micro-worlds are explicit representations listing all world objects with their properties, from which both the image and a caption is extracted. The objects are randomly sampled and, for now, consist of coloured shapes. For caption generation we use the Dependency Minimal Recursion Semantics (DMRS) formalism (Copestake et al., 2016; Copestake, 2009) to represent the abstract semantic structure of a proposition. Every object and property is annotated with its corresponding DMRS predicate(s), and the compositional framework of DMRS enables us to construct a wide variety of possible sentences from a few general DMRS graph skeletons on this basis. DMRS graphs can be transformed to MRS structures, from which corresponding English sentences can be generated with a bidirectional HPSG-grammar like the English Resource Grammar (Flickinger, 2000;Flickinger et al., 2014) and a parser-generator like ACE (http://sweaglesw.org/linguistics/ace/).

---

10. Grammatical constraints on lexical and structural processing strategies: EMEG evidence from Russian morphosyntax

Anastasia Klimovich-Gray, Elisabeth Fonteneau, William Marslen-Wilson
Dept. of Psychology

Sentential context can place strong constraints on the grammatical category of upcoming items and potentially the processing strategies they invoke. We examined the role and timing of such effects during the processing of derived/nominal and inflected/verbal forms. Inflections and derivations encode distinct grammatical and lexical information and are analysed via different processing strategies. Inflections carry grammatical information, accessed by decompositional processes that are related to left fronto-temporal activity. Derivations carry primarily lexical information, do not require decomposition and are primarily processed within bilateral temporal areas.

Using combined EEG and MEG we tested whether, and how early, the presence of grammatical constraints can activate a word's grammatical and lexical features, inducing a specific processing strategy. We recorded brain activity while subjects listened to Russian sentences with target words whose grammatical category (noun or verb) was either constrained or unconstrained by the preceding context. Nouns contained derivational suffixes, while verbs contained inflectional suffixes. In the unconstrained condition the target's form and category were ambiguous until the suffix onset, while in the constrained condition these were predictable. For example, the stem 'rabot-' can take an inflectional suffix to become a verb 'rabot-ayet'/'works' or a derivational suffix to become a noun 'rabot-nik'/'worker'. In 'Misha starij rabot-nik'/ 'Michael is an old worker' the stem is constrained by the preceding adjective 'starij'/'old' to be a noun while in ''Misha chasto rabot-ayet' / 'Michael often works', the adverb 'chasto' requires 'rabot-' to be a verb. In unconstrained contexts such as 'Misha rabot-nik na zavode'/'Misha (is a) worker at the factory' the class of the 'rabot-' is ambiguous until the affix is heard.

We hypothesised that in constrained conditions listeners predict the target form before the disambiguating suffix onset, thereby leading to different word-internal (inflection- or derivation-related) processing strategies. The results show that processes related to inflectional/verbal and derivational/nominal processes change as a function of preceding constraints. Unconstrained inflected forms produce distinctive left fronto-temporal activity with LIFG engagement peaking after suffix onset. This result is in line with previous findings, suggesting decompositional processing of inflected forms. In constrained contexts inflections are processed in the middle and anterior temporal areas, with no LIFG involvement. Derivational/nominal processing was less affected by these constraints and showed effects in the temporal areas after the suffix onset. This suggests that contextual predictability induced by grammatical constraints can facilitate lexical analysis of complex words and reduce the associated decompositional demands related to LIFG function.

---

11. Graph- and surface-level sentence chunking

Ewa Muszyńska, Computer Laboratory

Long sentences are a challenge for many Natural Language Processing (NLP) tasks, such as parsing. We propose sentence chunking as a new task reducing the complexity of sentences for further NLP processing. Chunking a sentence means cutting a complex sentence into grammatical constituents that can be processed independently and then recombined without loss of information. The task bears similarity to sentence simplification (Chandrasekar et al., 1996;Woodsend and Lapata, 2011) and clause splitting (Tjong et al., 2001).

Chunking can be defined both on the surface string of a sentence and on its semantic representation. We approach sentence chunking using rules defined on Dependency Minimal Recursion

Semantics graphs (DMRS; Copestake, 2009) and based on English Resource Grammar (ERG; Flickinger, 2000; Flickinger et al., 2014). We divide a semantic representation into subgraphs corresponding to logical chunks – the link structure of a DMRS graph reveals appropriate chunk boundaries.

Currently we perform chunking based on three types of syntactic structures: clausal coordination, subordinate conjuction and clausal verb complement. Based on our preliminary experiments, we report an improvement in the precision of chunking from 19.6% to 42% compared with simple string heuristics. Sentence chunking understood as graph chunking of a semantic representation can be directly useful for applications that already use the representation. Although we use the DMRS, chunking could be just as well adapted for other semantic representations, for example AMR (Abstract Meaning Representation) (Banarescu et al., 2013). We are also working on using the rule-based system to generate a dataset of sufficient quality to be used as training material for a minimally supervised machine learning system. Sentences could then be chunked without creating a full parse first and we could apply chunking to an extended set of NLP tasks, such as statistical machine translation or parsing itself.

---

12. Multi-level representations in speech processing in brain and machine: evidence from EMEG and RSA

Cai Wingfield (1), Li Su (2), Barry Devereux (1), Xunying Liu (3,4), Chao Zhang (3), Phil Woodland (3), Elisabeth Fonteneau (1), Andrew Thwaites (1), and William Marslen-Wilson (1).

1 Department of Psychology, University of Cambridge
2 Department of Psychiatry, University of Cambridge
3 Department of Engineering, University of Cambridge
4 Department of Systems Engineering and Engineering Management; Chinese University of Hong Kong

Human speech comprehension is an important test case for examining the relationship between the neurobiological systems supporting human cognition and emerging computational systems capable of emulating these capacities. Human listeners can accurately identify words in a continuous, noisy speech input, but there is only limited neuro-computational understanding of how this is achieved. Machine-based automatic speech recognisers (ASRs) are approaching human levels of accuracy. These systems are built strictly on engineering principles but provide a computationally specific model of successful speech-to-lexicon mapping.

Here, we ask whether ASR-derived models of speech recognition can be linked to the human solution. To do so, we relate dynamic brain states in the human listener to dynamic machine states in HTK, a candidate ASR based on a deep neural network (DNN) architecture. Simultaneous electro- and magnetoencephalography provides high-resolution recording of neural activity. Representational similarity analysis (RSA) makes it possible to compare brain and machine responses at the appropriate level of abstraction.

Low-level perceptual representations of speech are transformed in bilateral superior temporal cortex into a set of non-perceptual features, providing an intermediate representation of speech for mapping onto lexical representations. Here we use an ASR system to generate a comparable intermediate representation, and test the fit of this model to human brain responses in temporal cortex. We contrast this with a model capturing the basic acoustic-phonetic properties of speech.

The DNN reads mel filterbank (FBK) values at its input layer, and has five fully-connected 1000-node hidden layers (HL1–HL5), followed by a 26-node "bottleneck" layer (HL6), which compresses the speech representation to a low-dimensional code, read by a phonetic output layer. We used the FBK values to create models representing low-level acoustic properties of speech, and other layers to create models of intermediate-level representations. Each of these models was tested incrementally against brain data representations in bilateral superior temporal cortex. ASR-derived models each showed significant fit to neural response patterns in regions of left temporal cortex. These regions have frequency-selective and phonetic-feature-sensitive properties. Early hidden layer models (HL1–HL2) fit better than the purely acoustic model (FBK). Fit to later hidden layer models (HL3–HL5) dropped off, but again re-emerged for the bottleneck layer (HL6). We further analysed hidden-layer representations using a multidimensional scaling technique, and showed that articulatory features were descriptive of their general arrangement.

---

13. Neuro-computational modelling of lexico-syntactic representation and integration during speech comprehension

Hun S. Choi, Billi Randall, Barry J. Devereux & Lorraine K. Tyler
Centre for Speech, Language and the Brain, Dept. of Psychology, University of Cambridge

Speech comprehension engages complex cognitive processes, including the rapid activation of the lexical properties of words and their on-line integration into the developing sentence. As a result of developments in computational linguistics, larger-scale probabilistic models derived from corpora allow researchers to unveil the spatio-temporal dynamics of incremental speech comprehension. We examined the relationship

between the detailed lexico-syntactic information activated when a word is heard, and processes of integration by investigating how strongly the lexical constraints of a verb influence the expectation of an upcoming complement structure in sentences such as:

(a) The young man fled the scene of the terrible accident (DO complement, frequent for 'fled')
(b) The young man fled to the forest when the chase began (PP complement, less frequent for 'fled')

In the sentences, the surprisal (i.e., how expected that structure is) of the given complement structure varied as a function of the lexico-syntactic properties of the preceding verb, providing sufficient variability to detect ease/difficulty of integration of the complement structure. We derived corpus-based cognitive models to test against multivariate neural activity patterns recorded by combined electro- and magneto-encephalography (EMEG). We hypothesized that lexico-syntactic representation modeled by the relative frequency of possible complement structures (or subcategorisation frames (SCFs)) would recruit left posterior middle temporal gyrus (LpMTG) and left inferior frontal gyrus (LIFG) soon after verb onset (see Tyler & Marslen-Wilson, 2008). The process of syntactic integration was modeled by the Bayesian likelihood (or backward conditional probability: P(v|cs) where v = verb and cs = complement structure) and was hypothesized to be located in LIFG soon after the complement word (e.g. 'the'/'to' in (a)/(b)) is recognized (Tyler et al., 2013). Here, we adopted a multivariate pattern modeling approach mainly because comparing pattern matrices rather than amplitude vectors allowed us to effectively combine information varying over multiple dimensions (e.g. SCFs; spatio-temporally varying neural activity). Our analysis is particularly suited for multi-dimensional modeling and, hence, to our question regarding 'spatio-temporal dynamics' that conventional event-related-potential/field (ERP/ERF) studies cannot address.

Consistent with our hypothesis, we found a significant lexico-syntactic representation in a left-lateralised fronto-temporal language network, starting at 200ms from verb onset in LpMTG and gradually moving into LIFG. Furthermore, a significant correlation between LIFG and our syntactic integration model was observed around 300ms after the complement word onset (as phonological processing was completed in left Heschl's Gyrus (LHG)). These results elucidate how incremental processes evolve, transition and decline over space and time in the brain. Moreover, our study corroborates the role of left fronto-temporal network in syntactic processing and distinguishes between activation of lexico-syntactic knowledge and syntactic integration.

---

14. Overnight changes in the neural representation of newly-acquired inflectional affixes: behavioural, univariate, and multivariate fMRI evidence

Viñals, L. (1, 2) , Mirković, J. (3, 4), Gaskell, M. G. (3), Davis, M. H. (1)

1 MRC Cognition and Brain Sciences Unit
2 Dept. of Theoretical and Applied Linguistics
3 University of York

Inflectional morphology is a proving ground for investigating whether learners generalise based on how often individual words occur in the input (token frequency) or based on how many different words in the input follow a particular pattern (type frequency). In two experiments, we focused on the role of overnight memory consolidation in influencing the representation and generalisation of novel inflectional affixes trained with different type and token frequencies.

Over the course of 9 days, we trained participants on an artificial plural system in which novel words referred to the occupation of male and female characters. On Day 1, participants learned the singulars (e.g. *gleeti* [fem, sing] = doctor [fem, sing], *shilnu* [masc, sing] = painter [masc, sing]). On Day 8, participants learned a first set of plural affixes for one gender (e.g. *gleetaff* [fem, plur] = doctors [fem, plur]). On Day 9, participants learned a new set of plural affixes for the other gender (e.g. *shilnopp* [masc, plur] = painters [masc, plur]). This design allowed us to make within-subject/between-affix comparisons in the generalisation and representation of affixes trained on consecutive days. The novel words were designed such that the majority were phonologically varied and took a high type frequency regular affix (e.g. *gleetaff* [fem, plur], *shilnopp*[masc, plur]), akin to phonologically varied regular English verbs (*call-called, turn-turned*). A subset of words contained an ambiguous phonological cue (e.g. arb) which was associated with both a high token frequency irregular affix (e.g. *varbesh* [fem, plur], *yarbull* [masc, plur]) and a high type frequency regular affix (e.g. *farbaff* [fem, plur], *tarbopp* [masc, plur] but also *gleetaff*[fem, plur], *shilnopp*[masc, plur], etc.). This mimicked phonological similarities between regular and irregular English verbs (*e.g. fit-fitted, sit-sat*).

In Experiment 1, productive generalisations of the plural affixes to untrained phonologically ambiguous singulars (e.g. *zarbi* [fem, sing], *zarbu* [masc, sing]) showed greater influence of token frequency for affixes trained on Day 8 (consolidated) than for affixes trained on Day 9 (unconsolidated). In Experiment 2, we examined this consolidation effect further by using fMRI to compare neural responses to affixes trained on Day 8 (consolidated) and Day 9 (unconsolidated). Combined univariate and multivariate analyses revealed consolidation-related changes in the representations of the trained affixes, modulated by training frequency, in the left superior temporal gyrus, the left inferior frontal gyrus, and the left posterior

hippocampus. We discuss these findings with reference to a Complementary Learning Systems account of learning and memory.

---

15. Pitch-interval analysis of 'periodic' and 'aperiodic' Question + Answer pairs

Juan Pablo Robledo (1) , Sarah Hawkins (1) , Ian Cross (1) , Richard Ogden (2)

1 Centre for Music and Science, University of Cambridge
2 Dept. of Language and Linguistic Science, University of York

Recent research shows that, in English Question + Answer (Q+A) pairs, periodicity typically emerges across turn space, to a degree of precision matching standards of music perception. Interactionally-aligned Q+A pairs display such shared periodicity across the turn, while unaligned pairs do not. Periodicity is measured as the temporal location of f0 maxima or minima in successive accented syllables (following Loehr's (2007) terminology, "pikes") within and across turns. The present study asks whether periodicity of pikes across a turn is accompanied by systematic use of musical pitch intervals across the turn space (regardless of language-specific intonation patterns). The sample consists of recordings of 77 Q+A pairs from eight same-sex pairs of friends (four pairs each sex, aged 19-31 years, mean=24) who were asked to talk naturally as well as jointly improvise music. All were university educated, native speakers of British English (England, Scotland, Northern Ireland). Ratios of f0 in the last pike of the Question and the first of the Answer fell more reliably into Western musical interval categories when the Q+A pair's turn transition was periodic (the Answer was aligned or preferred, re the Question) than when it was aperiodic (disaligned, dispreferred). Similar results were found for ratios of modal f0. Such pitch ratios are better described by musical interval categories of Western tuning systems than by those of three non-Western systems, and best of all by semitones, suggesting close connections between culturally-specific uses of pitch in conversation and in music. Theoretical implications are discussed.

---

16. Sublexical morpheme stripping

Luca Cilibrasi, Dept. of Theoretical & Applied Linguistics

A strongly debated issue in psycholinguistic research regards the way in which inflected verbs are processed. Do we process stems and bound morphemes as units or do we separate stems and bound morphemes? For instance, when we hear the sentence "she played football yesterday", do we process "played" as a unit, or do we decompose it in "play" + "ed". Experimental evidence seems to be contradictory, confirming one or the other hypothesis (Stemberger & MacWhinney, 1986, Pinker & Ullman, 2002), although there is relative agreement that the answer may strictly depend on the frequency of the verb (Caselli et al. , 2016). Our project is investigating this problem focusing on sublexical items, in order to avoid the confounds generated by the access to the lexicon.

In our study, participants are presented with pairs of nonwords that contain or do not contain potential morphosyntactic information. The experiment is based on a peculiar morphophonological property of English. In English, regular verbs ending in /l/ take /d/ as bound morpheme when inflected in the past and /z/ when inflected at the third person singular. For instance, the verb "to kill" becomes "killed" in the past and "kills", pronounced /kilz/ in the present. The unvoiced relatives of /d/ and /z/ are /t/ and /s/. /t/ and /s/ can have a morphosyntactic meaning, for instance when following a verb ending in /k/. The past tense of the verb "bark" is an example: it becomes "barked" in the past, pronounced /barkt/ and "barks" at the third person present. Crucially, /t/ and /s/ cannot have a morphosyntactic value when following a verb ending in /l/.

Given this, 40 nonwords ending in /l/ were created, and they were combined with /z/ and /d/ as endings, leading to potentially morphosyntactic minimal pairs, or with /t/ and /s/, leading to non-morphosyntactic minimal pairs. A further condition with 40 items ending in /b/ and /m/ was added, to control for voicing effects. The perception of these three types of minimal pairs was investigated with a same-different minimal pairs discrimination task, conducted on adult native speakers. The results show that elements in a potentially morphosyntactic minimal pair took longer to be discriminated than elements in a non-morphosyntactic minimal pair. The result is not likely to be due to voicing difference between the two conditions, since morphosyntactic minimal pairs required more time (marginally significant difference) also when compared to the voicing control condition. This result suggests that, at least, subjects are subconsciously processing differently the presence and the absence of potentially morphosyntactic information. Further, it shows that this discrimination is not blind to the nature of the stem, otherwise pairs ending in /t/ and /s/ would not be different from /z/ and /d/.

In the next two years the study will be extended to children with and without dyslexia, and to a group of bilingual children.

---

17. The development of svarabhakti vowels in the history of Norwegian
Tam Blaxter, Dept. of Theoretical & Applied Linguistics

Old Norse allowed a number of marked consonant clusters which are no longer allowed in any of the modern North Germanic languages; most frequent are coda clusters in /Cr/, as in *vetr/vintr* 'winter' and *akr* 'field'. These consonant clusters were removed by the introduction of svarabhakti vowels in the Middle Ages. Thus from *vetr/vintr* we find Modern Icelandic/Faroese *vetur*, Norwegian *vetter* and Danish/Norwegian/Swedish *vinter* (Magnússon 1989:1128); from *akr* Icelandic/Faroese *akur*, Swedish/Norwegian *åker* and Danish *ager* (Magnússon 1989:9). The change presumably involved two steps: the insertion of phonetic svarabhakti vowels to break up these clusters without any change in underlying forms and the phonologisation of these new vowels. We can assume that only the latter change will be visible in written records. Although the rise of svarabhakti vowels took place in all North Germanic languages, there are differences in the progression of the change. The vowel quality varies between languages: it was originally /u/ in Icelandic and Faroese, schwa in Danish, and variable within Norwegian and Swedish. The change in writing took place at different times in the different North Germanic languages: already in the earliest written records in Danish; in the early fourteenth century in Swedish and Norwegian(Indrebø 1951:118); and in the sixteenth century in Icelandic.

This research investigates the change in Old and Middle Norwegian. Using medieval charters dated, localised and categorised for social factors, it aims to determine the timing of the change more precisely than has been possible in previous research. It aims to identify the place of origin of the change in Norway and describe the process of diffusion by which it spread to all Norwegian speaking areas: as the change took place earlier in Danish than in Norwegian, it might be expected that the change spread into Norway through Sweden from Denmark; this paper will test this hypothesis. It also aims to elucidate the social progression of the change. If contact with Danish was responsible for the spread of this feature in Norwegian, it can be predicted that the social groups most subject to contact with Danish (merchants, officials) would lead the change. Finally, the distribution of the different qualities of the svarabhakti vowel within Norway are investigated and mapped.

---

18. The effect of physical activity on implicit language learning in children and adults
Carla Pastorino Campos, Dept. of Theoretical & Applied Linguistics

Research linking physical activity and cognitive task performance has shown that chronic and acute exercise may have a positive effect on cognition in general, and learning in particular (Chaddock, 2012; Coles & Tomporowski, 2008; Etnier et al., 2014; Labban & Etnier, 2011; Pesce et al., 2009; Schmidt-Kassow et al., 2010; Winter et al., 2007). While research efforts have mainly concentrated on learning activities involving declarative memory, few studies have investigated the possible link between exercise and procedural memory. Procedural memory has been posited to play an important role in language learning, as it is thought to be involved in the learning of rule-like regularities (Ullman, 2004). To address this gap in the research, this project seeks to explore the possible effects of acute physical activity on the incidental learning of higher-level, linguistic-based regularities, a type of learning considered as relying on procedural and implicit memory systems. In addition, it intends to widen our understanding of how physical activity may influence cognition across the lifespan by using the same language learning activity with children and adult participants.

Groups of highly fit children and adults participated in an implicit rule-learning task (Leung & Williams, 2014) presented as a computerized game. They were exposed to pairs of known words together with novel non-words. They were taught that half of the novel non-words meant 'near' and the other half meant 'far', but were unaware that one set of near/far words occurred only with animate nouns while the other only with inanimate. During exposure, the children saw novel-known word pairings and were asked to indicate the animacy of the known word and the distance of the novel word. The responses were made using buttons in a keyboard while reaction times were recorded. The testing phase was identical to the exposure, but the pairings now included both trials that followed the hidden rule and trials that did not. This language learning activity was preceded either by a period of aerobic physical activity or rest.

Preliminary analyses of the data show that for these adults a bout of exercise did not have an effect on their performance on the rule-learning task. For children, those who exercised did not seem to learn the hidden rule. This finding could indicate that while previous research has shown that exercising prior to learning positively affects learning in declarative tasks, this advantage may not only not translate to other memory domains but might even reverse in child populations.

---

19. The effects of attention and interference on the neural encoding of continuous speech
Andrea Olguin, Tristan A. Bekinschtein, Mirjana Bozic

Dept. of Psychology

Knowing two languages is thought to lead to an enhancement in selective attention in bilinguals (e.g., Bak et al., 2014; Krizman et al., 2012), however this remains a hotly debated issue (Paap et al., 2015). It is also unclear what neural mechanisms would support this phenomenon. This project aims to investigate these issues by examining the neural encoding of attended continuous speech in early bilinguals, late bilinguals, and monolinguals. We will also investigate how different types of interference influence the neural encoding of attended speech. In this first study, we examined how attention modulates the neural encoding of continuous speech in monolinguals, aiming to provide a baseline against which the effects in bilinguals can be assessed. In a dichotic-listening task, English monolinguals attended to a narrative in English presented to one ear, while ignoring interference presented to the other ear. Four different types of interference were presented to the unattended ear: a different English narrative, a narrative in a language unknown to the listener, a well-matched non-linguistic acoustic interference (Musical Rain); and no interference. Neural encoding to attended and unattended speech was measured by calculating cross-correlations between the attended and unattended speech envelopes and the neural activity as recorded by a dense array 128-channel EEG system. Findings show that there is a more robust neural encoding for the attended envelopes compared to the ignored ones. We also observed differences across the four conditions, indicating that the type of interference modulates the neural encoding of attended continuous speech. There was significantly greater encoding of attended speech in the no interference condition, relative to attended speech in the presence of either linguistic or non-linguistic interference. Within the three interference conditions, the strongest encoding of the attended speech was found when the interference was in a known language (English) and weakest when the interference was non-linguistic noise (Musical Rain). There were no differences in the encoding of the unattended streams. Overall, these results show that attention and type of interference modulate how speech is encoded in the brain, and suggest that the enhancement of the attended stream (rather than the suppression of the unattended one) might play a key role here. The follow up experiments will compare these effects with the equivalent processes in bilingual listeners.

---

20. The interaction of syntactic structure and lexical constraints during sentence processing

Barry Devereux, Billi Randall, William Marslen-Wilson, Lorraine K. Tyler
Centre for Speech, Language and the Brain, Dept. of Psychology

Understanding a spoken sentence involves a complex set of processes that integrate the stream of words into a coherent representation of the utterance's structure and meaning. A number of factors have been proposed to influence the ease with which a word can be integrated into the unfolding representation, including the overall syntactic complexity of the sentence (Friederici, 2002; Gibson 1998) and probabilistic lexical knowledge about the kinds of structures a word tends to be used in (MacDonald et al, 1994; Marslen-Wilson, 1973). Reflecting these separate accounts, neuroimaging studies investigating integrative processes in sentence comprehension have often either manipulated syntactic complexity whilst strictly controlling or eliminating lexical influences, or have manipulated lexical properties in the context of simple grammatical structure. In the current MEG study, we explore how lexically-driven expectations and syntactic complexity interact during the incremental interpretation of spoken sentences by building corpus-derived models of lexical information and investigating their influence on long-distance structure building. Seventeen participants listened to sentences where an intervening clause could separate the subject and main verb of the sentence, creating a long-distance dependency. We included three long-distance dependency conditions: (a) sentences containing a central phrase that was unambiguously a relative clause (e.g. "The van *that was clamped on the driveway* had run out of petrol"); (b) sentences containing a reduced relative clause with a highly transitive verb (e.g. "The van *clamped on the driveway* had run out of petrol") and (c) sentences containing a reduced relative clause where the verb could be used intransitively ("The van *stalled on the driveway* had run out of petrol"). Conditions (b) and (c) differ only because of the different lexico-syntactic expectations associated with the verbs ("clamped" vs "stalled") – the prepositional phrase in (b) indicates a passive construction with "clamped on the driveway" functioning as a relative clause, whilst the prepositional phrase in (c) can be interpreted as an adjunct to the verb. This creates a classic garden path in (c), where the main verb ("had") is difficult to integrate with the preceding context. In the source-space MEG RSA analysis, we found significant sensitivity to the surprising reduced-relative condition (c), compared with the unsurprising conditions (a, b), in bilateral STG and left MTG. This effect was predicted by a probabilistic corpus-derived measure of verb direct object preference, with greater direct object preference corresponding to lower surprisal. These results are important in demonstrating how specific lexical knowledge about verbs – modelled using corpus data – interacts with complex syntactic processing during spoken language comprehension, yielding a detailed picture of processes of integration in sentence processing.

---

21. The neuromagnetic time-course of semantic ambiguity resolution in speech comprehension

Lucy J MacGregor (1), Jennifer M Rodd (2), Olaf Hauk (1), Matthew H Davis (1)

1. MRC Cognition and Brain Sciences Unit, Cambridge
2. University College London

Semantically ambiguous words challenge sentence comprehension, particularly when the disambiguating context is delayed and serves to select a non-dominant meaning making reinterpretation necessary. We investigated neural responses associated with meaning selection and reinterpretation during semantic ambiguity resolution using MEG/EEG. Volunteers listened to spoken sentences containing ambiguous words and delayed *disambiguation* (e.g., *Sally worried that the BALL was going to be too crowded*). These sentences engage selection/reinterpretation processes compared to sentences: (1) containing unambiguous control words (e.g. *PUB* substituted for *BALL*) or (2) in which the final word no longer favours the non-dominant meaning (*expensive*, not *crowded*). Listeners heard four variants of 80 such sentences in a two-by-two factorial design allowing us to assess responses associated with initial meaning activation and subsequent reinterpretation of ambiguous words. Greater activity for ambiguous compared to control words was observed over left fronto-temporal regions 420-800 msec after word offset. This response correlated positively with individual differences in comprehension which was predicted by participants' vocabulary scores. Reinterpretation was not associated with additional neural responses, but a measure of the surprisal of sentence-final words was negatively correlated with comprehension and with activity over right fronto-temporal regions. Implications for the neurocognitive mechanisms of ambiguity resolution will be discussed.