

	A	S	T	O	N		
	C						
C	O	R	P	U	S		
	R						
	N	E	T	W	O	R	K

Pre-ACORN (1)

- **No Data (NB: refs to ASTEC Aston Scientific and Technical English Corpus, etc)**
- **ATA (Aston Text Analysis) Software**; frequency, concordance, collocation by position; Mark 1, MS-DOS (16-bit); Mark 2 for Windows 95, NT4, or later
- **Software Problems** (eg distance MSc students); PCs only; filesize and format constraints, punctuation (fullstop, quotation marks, hyphen rows, etc); slow (5-6 minutes to create index for 0.25m words; research tool, not learner interface; mainly intended for English, no choice of language normally available; software support by Peter Roe (retired); commissioned from Forwords Limited (Lichfield?) by LSU (so changes difficult)

Pre-ACORN (2)

- **Ramesh** (Aug 2004); AppleMac; personal access to BoE; downloaded data sporadically; Wordsmith Tools
- **Chris** (July 2005); software development; Ramesh and Emmanuelle to TWC
- **FLDC** call for bids (May 2005); proposal (Pam); negotiations (e-learning; Beverly and Nic Meara); **Sylwia** and **Stefan**; Emmanuelle; Staff Questionnaires
- **FLDC** funds arrived (Dec 2005); **ACORN** project officially started (Jan 2006)

ACORN

(Aston Corpus Network)

- **Stage 1: January 2006 – January 2007**
- **Funded by the Flexible Learning Development Centre, Aston University**
- **AIMS:**
- **Corpora in English, French, German and Spanish (5-10m words, incl. specialized corpora); translation corpora (1m words per language; originals and translations); academic corpora: student texts; textbooks, research articles**
- **Software to process and analyse the data**
- **Pedagogically-oriented outputs (incl exercises, testing)**
- **Dissemination**
- **Training (staff and students)**

ACORN Project Team

Ramesh Krishnamurthy (English; Aston)

Sylwia Jaworska (German; Aston)

Stefan Baumgarten (Translation; Aston)

Iztok Kosem (Academic; French; Spanish; Aston)

Christopher Martin (Programming; Aston)

Constantin Orasan (Consultant; Wolverhampton)

Emmanuelle Labeau (Associate; French; Aston)

*new appt in Sp/linguistics (Associate; Spanish; Aston)

Branka Visnjic (Honorary; Pedagogic; City College)

ACORN progress: Jan-May 2006

- **Staff Questionnaires**
- **Identifying/Locating Sources**
- **Obtaining Permission**
- **Acquiring/Downloading Data**
- **Monolingual Corpora: *English*: Enron emails (3m); Brown (1m); WBE (10m); Wton-junk-emails (0.3m); MSc-TESOL (?); *German*: Spiegel-Online + Archive; *Spanish*: Gutenberg and other texts;**
- **Translation Corpora: *English-French*: Human Rights; Madame Bovary; *English-German*: Springer Medical Abstracts; *English-French-German-Spanish*: EU Constitution; *English-Spanish*: in progress**
- **Hardware: Development machine (70GB), Webserver (30GB), Central filestore (2GB)**

ACADEMIC SUB-CORPUS

- **Obtaining material from students:** how?, when?; the importance of lecturers' endorsement
- **Initial stage:** analysis of academic written language, EAP dictionary.
- **Future aims:** playing part in curriculum design, teaching material design, pre-sessional teaching, exercises.

SOFTWARE DEMONSTRATION

- **Corpus Selection – language, variety, genre, vintage**
- **Word Frequency lists – the (relative) importance of an item**
- **Ngram Frequency lists – the exact common phraseologies of an item**
- **Concordances – the co(n)texts/texts an item is used in**
- **Concordances – sorting: the close lexical and syntactic company of an item (collocation and colligation)**

ACORN: the future

- **User trials, user interface design, output design; documentation; training**
- **Security backup procedures**
- **Complete Stage 1 (by Jan 2007)**
- **Funding for Stage 2 (2007)**
- **Feedback from staff and student users**
- **Software enhancements (eg collocation)**
- **Stand-alone version of software to run on user data**
- **Automation of data collection**
- **Corpora enhancements: more data, more variety, more specialised**