Informe of Programming Languages

Proposer(s): James Cheney
Date: October 20, 2014
Status: DRAFT for discussion at Board of Studies

1 Case for Support

1.1 Overall contribution to teaching portfolio

Programming language design and implementation (PL) is a broad research area that interacts with many other areas of computer science, AI and software engineering. It is ranks third largest among core areas in the ACM CS2013 curriculum (and second in terms of advanced “tier 2” material) and corresponds to the “Comparative programming languages” area of the the QAA Subject Benchmark for Computing.

![ACM CS2013 curriculum summary](image)

Figure 1: Summary of ACM CS2013 curriculum tier1 and tier2 hours, by subject area

The proposer carried out a curriculum analysis of PL-related courses earlier this year to match our existing offerings against the ACM guidelines (this analysis is attached to this proposal). Currently only 12 of the ACM-recommended 28 hours of core PL material are covered in required courses such as INF1-FP, INF1-OOP and INF2A, and an additional 5-6 hours of core material is covered in optional honours courses such as CT and ITCS. This leaves roughly 10 hours of core material uncovered by any Informatics course, and that is the best case: it assumes that students take both CT and ITCS (currently, only seven students are enrolled in both courses in Spring 2015). Thus, the vast majority of our students currently only see 12–15 of the 28 recommended hours of material.

Although it might (at first glance) seem easy to fill this gap merely by adding an extra lecture or two to several existing courses, this would not really address the problem. This material is mutually reinforcing so simply adding piecemeal coverage to other courses would be far less effective than presenting it as part of a unified topic. Moreover, even if material were added to current PL-related courses to fill this gap, students would still not see this material unless they took all of these courses; this is unlikely because the more specialised UG4 electives are typically taken by under 30 students.

Therefore, we propose to develop a new course that covers the 10 hours of material that our students currently do not have any opportunity to learn. To round out the course, the remaining approximately 5 hours of examinable material can be designed to cover the ACM curriculum areas already covered by other optional courses CT and ITCS (from a different perspective, and without duplicating material.) The course will emphasise practical issues, complementing our existing slate of more theoretical or advanced courses, and will be designed to appeal to students who might not otherwise take CT, ITCS, or more advanced PL-related courses.

Previous courses: LSI and FPS used to cover this area but are no longer being taught, in part due to declining demand. There is no evidence that the decline in demand for those two courses
is due to decline in interest in PL topics more generally; other PL-related courses such as CT, PPLS, APL, LP, COC, TSPL, or SP — some of which used to depend on LSI or FPS as prerequisites — have not seen a comparable decrease in demand.

**Demand:** New programming languages (including scripting languages and domain-specific languages) are appearing all the time, and we cannot (and probably should not try) to teach all of them. Instead, teaching students the underlying ideas of programming language design will prepare them to learn new languages more quickly and effectively; this does not just mean knowing how to use existing features correctly but also what features to avoid and why their use is problematic.

Although few CS graduates will ever design a new general-purpose language, many will be faced with the task of learning, adapting or designing a new scripting language or domain-specific language. Currently, many such languages are not consciously designed but instead grow from a custom tool to meet some need; this usually involves reinventing known features, and sometimes results in re-making known mistakes. It is important for computer science professionals to be aware of programming language design as a coherent discipline, so that they know where to look for ideas if they are faced with a similar task.

PL concepts arise in many research areas in the School, and familiarity with techniques such as abstract syntax, interpreters, domain specific languages, translators, and types would contribute to potential UG or MSc projects in several areas beyond core PL, such as databases (query languages), natural language processing, security, networks, and quantum computation.

**Competition:** Peer UK and international institutions all offer courses on PL concepts, for example:

- Programming Languages, Stanford University, [https://courseware.stanford.edu/pg/courses/317431/cs242-fall-2012](https://courseware.stanford.edu/pg/courses/317431/cs242-fall-2012)

and some of these (e.g. CMU’s 15-312) are listed as exemplars in the ACM CS curriculum.

The University of Edinburgh is known worldwide for fundamental contributions to the foundations of programming languages, including many topics that are now routinely taught in the courses listed above. We can and should offer one of the world’s best courses on programming language design to our own students, whether or not they go on to pursue more advanced topics offered by our UG4 courses on PL.

### 1.2 Target audience and expected demand

**Background:** UG3 or UG4 students having taken Inf1-FP, Inf1-OOP, Inf2A.

**Level of ability:** Medium to high. The course would be demanding.

**Interests of students likely to take this class:** Programming, programming languages, symbolic computation, compilers, interpreters.

**Expected class size:** This course would likely appeal to students that currently take CT, PPLS, COC, TSPL, APL, LP, AR, and ITCS, as well as to students who might be interested in learning about programming languages but might not feel prepared to take the more advanced courses we currently offer. The enrollment figures for these courses for the current and previous academic year are:
where italics indicate enrollment figures for Spring 2015 courses that are not finalized yet. Based on comparable UG3 courses such as CT, LP, CT, and ITCS, it seems reasonable to anticipate initial class size in the range 15–25 students. This could increase if the course design is successful in attracting students who might not naturally have taken other upper-level courses on this topic, and may have the (desirable) side-effect of increasing demand for these courses or for honours projects in this area.

1.3 Relation to existing curriculum

Fit with existing courses The course is tentatively planned to be taught in semester 1, normally to be taken in UG3, but also open to UG4 or suitably interested MSc students. The following diagram indicates how it fits into the current curriculum:

In more detail:

- EPL will **build upon and consolidate** knowledge of functional programming, object-oriented programming, and formal languages established in INF1-FP, INF1-OOP and INF2A.

- EPL will **complement** other Honours courses normally taken in UG3 related to programming languages, such as CT, LP, and ITCS. Specifically:
  - Compiling Techniques focuses on parsing, compilation and low-level code generation, and EPL will complement this with understanding of interpreters, translation and high-level language features
  - Logic Programming focuses on Prolog and logic as a foundation for programming, while EPL will complement this with broader understanding of programming language design ideas (including declarative/functional/logic programming, object-oriented programming, etc.)
  - Introduction to the Theory of Computer Science covers lambda calculus and types in some depth from a theoretical perspective, including topics such as type inference that
may be covered more broadly and less formally EPL. EPL will complement this material by presenting pragmatics about language design, interpretation and translation. The two courses would remain independent, and could be taken in either order.

- EPL will **strengthen students’ preparation for advanced PL courses** normally taken in UG4, such as APL, TSPL and COC, and may increase demand for these courses or motivate additional offerings at UG4 or MSc level.

- EPL will also **strengthen students’ preparation for PL-related UG4 projects** involving language design or implementation as a component or main focus.

- EPL may also be useful preparation for UG4 courses such as SP, PPLS and COPT, but this is not an explicit design goal, since the progression of students from CT and CS in UG3 to these UG4 courses appears to be healthy already.

- Likewise, EPL would complement AR. AR is currently taught in UG4 but I am aware of plans to move it to UG3 and add a fourth year Formal Verification course. One important application area for AR and formal verification techniques is proving properties of programs or programming languages, and EPL would provide useful background for such material as well as potentially attracting students to the area. Students taking both EPL and AR will be well-prepared for the new Formal Verification course as well as the advanced course TSPL, in which programming language concepts are formalized in a theorem prover.

**Degrees and specialisms** I expect this course to contribute to:

- All undergraduate Degree Programmes spanning Computer Science and Software Engineering

- The MSc specialisms of Computer Systems, Software Engineering & High-Performance Computing, Design Informatics and Theoretical Computer Science. (It is relevant to Design Informatics in that programming language design often involves some aspects of aesthetics as well as technological problem-solving.)

This is based on the fact that the course covers core computer science concepts.

### 1.4 Resources

- **Lectures**: 18-20 per semester.

- **Tutoring**: 10 hours (one 2-hour introductory lab session + 8 smaller group tutorials)

- **Exam preparation**: As for a typical course taught in the autumn term. Anticipating the course would be open to visiting undergraduates, an extra December exam would need to be prepared whenever visiting undergraduates take the course.

- **Course organization team**: 1-2 lecturers + 1-2 tutors/markers. In its first year it is anticipated that the course would be delivered by the proposer, to ensure consistency in its development. The proposer is currently supported by a fellowship until December 2016, but is willing to voluntarily accept this teaching load.

- **Alternative lecturers**: Because this material is close to the research interests of several Informatics teaching staff, it is expected that it would be straightforward to find alternative lecturers.
2 Course descriptor

Course Title: Elements of Programming Languages

SCQF Credit Points: 10

SCQF Credit Level: 10

Normal Year Taken: 3

Also available in years: 4/MSc

Subject Area and Specialism Classification:

- Any BSc or MInf degree programme including the words “Computer Science” or “Software Engineering” or “Informatics”.
- MSc specialisms:
  - Computer Systems, Software Engineering and High Performance Computing
  - Design Informatics
  - Theoretical Computer Science

Appropriate/Important for the Following Degree Programmes: This course covers core material recommended for all students taking a Computer Science, Software Engineering or Informatics degree.

Timetabling Information: EPL is likely to be of interest to similar students to those taking UG3 level 9/10 courses such as LP, CT, AR, ITCS, and may also be taken in the fourth year by students also interested in courses such as APL, PPLS, COPT, AR, or TSPL. It would be helpful to avoid scheduling conflicts with these courses. For this reason it is proposed that EPL be taught in Semester 1, to avoid undue scheduling pressure, because the majority of these complementary courses (CT, AR, ITCS, PPLS, COPT, TSPL) are taught in Semester 2.

School Acronym: INF-?-???

Short Course Description: Programming languages are unique forms of communication that play a dual role: not only as ways for programmers to instruct machines, but as ways for programmers to talk to each other about computation. Paradoxically, they are among the most permanent features of the computing landscape (Fortran, for example, is still widely used 60 years since its invention), and among the most energetic and innovative, with new programming languages introduced every few months, often aiming to simplify Web programming, parallel, or distributed computing.

Although few computer scientists will ever design a new, general-purpose language like Java or C++, all computer scientists need the ability to learn new languages quickly, recognise and use (or avoid misusing) common language features, and even design new domain-specific languages for restricted problem domains. The design of programming languages involves many subtle choices and tradeoffs among performance, convenience, and elegance. In this course, you will not learn how to use any one language, but instead you will learn the basic elements you need to understand the next 700 programming languages, or even design your own.
Pre-Requisite Courses: None (beyond required INF1 and INF2)

Co-Requisite Courses: None

Prohibited Combinations: None

Other Requirements:

- Undergraduate: Successful completion of year 2 of Informatics Single or Combined Honours degree, or equivalent by permission of the School
- Postgraduate: Available to postgraduates with proven programming ability (e.g., first degree in Computer Science, Software Engineering or similar)

Available to Visiting Students: Yes

Summary of Intended Learning Outcomes: A student who has successfully completed this course should be able to:

1. investigate the design and behaviour of programming languages by studying implementations in an interpreter;
2. employ abstract syntax and inference rules to understand and compare programming language features;
3. design and implement a domain-specific language capturing a problem domain;
4. understand the design space of programming languages, including common elements of current languages and how they are combined to construct language designs;
5. critically evaluate the programming languages in current use;
6. acquire and use language features quickly, recognise problematic programming language features, and avoid their (mis)use.

2.1 Assessment Information

Assessment Weightings: Written Examination: 75%
Coursework: 25%

Time spent on assignments: Three assignments taking roughly 5–10 hours each, for a total of 25 hours.

2.2 Academic description

The course first covers general considerations and fundamental concepts used throughout the rest of the course: the idea of abstract syntax and the use of inference rules to model the design of language features. This is expected to cover roughly the first five lectures. The remainder of the course builds on this foundation to explore the elements of programming languages, including data structures and abstractions, control structures and abstractions, and approaches to modularity and program organisation. This exploration is structured as a series of case studies, each presenting a programming problem, and exploring ways in which it has been addressed, ranging
from (sometimes naive) early approaches to those taken in modern languages. The intention is not just to denigrate what the instructors believe to be bad design, but to highlight design “mistakes” where it is clear that a particular feature (or combination of features) is prone to misuse, while offering a nuanced view of the many tradeoffs and interactions between reasonable design choices where there is no single “best” way.

For example, Hoare once called null pointers the “billion dollar mistake”. It is common to need a way to represent missing data in a program, but null pointers / null references accommodate this in a way that is particularly hospitable to subtle bugs. This issue reappears in many languages and has led to a large number of crashes and security vulnerabilities causing economic and other forms of damage. This could form the basis of a lecture contrasting the conventional approach to missing data with alternatives such as option types in functional languages, or even NULL values in SQL.

A single programming language will be used as a running case study of language design, with ideas from other languages introduced in lectures as necessary. The same language will be used for the coursework (programming) assignments; a two hour lab session will introduce students to the language used for the course and eight weekly tutorials will cover practical programming exercises or mini-case studies.

Exposure to features arising in a variety of programming languages is important part of the course, however, due to the length of the course it is not judged to be feasible for students to learn two or more unfamiliar languages fluently. Instead it is proposed to select a single programming language to be used for the coursework assignments. Since the purpose of the course is to learn the elements of programming language design, it makes sense to ask the students to learn a non-mainstream language that has a number of interesting features that can be used to illustrate good (or critique questionable) design, as long it is clear we are not endorsing this (or any current) programming language as an ideal. It would also be strategically reasonable to choose a language with some current credibility in industry, though the course will not teach programming in that language so much as use it as a primary case study of a complete language design.

Languages arguably currently meeting the above criteria include Scala, F#, Python and Ruby, Scheme/Racket, OCaml, Haskell or Standard ML would also be sensible choices but are less credible beyond academia. Of these, Scala seems to offer the best tradeoff between familiarity (building on the Java ecosystem), interestingness of design (functional programming, polymorphism, and sophisticated object-oriented features such as traits), and industrial relevance. For the sake of concreteness we propose that the course initially use Scala, but the course is not dependent on the particular programming language — in principle, the coursework and tutorials could be adapted to use any other language.

The course will feature a distinctive practical component aimed at establishing a baseline of literacy in language-oriented problem solving techniques (consolidating and integrating initial exposure to many such ideas in INF1-FP, INF1-OOP and INF2A). There will be three coursework assignments:

- a “warm-up” exercise to (re)familiarize students with functional programming and ensure that they are prepared for the more rigorous programming exercises to come;
- a coursework focusing on implementing an interpreter for a simple language and extending it with some non-standard features;
- a coursework focusing on the design and implementation of a domain specific language (e.g. translating high-level geometric primitives to triangles for input to a ray tracer, or simple markup/template language to HTML or PDF)

**Syllabus:** (not necessarily to be covered in order)
• Abstract syntax and name-binding; Lexical vs. dynamic scope
• Modelling programming language features using inference rules
• Interpreters as a way of defining and exploring language features
• Domain specific languages, translators, and pragmatics of language processing
• A selection of topics illustrated by case studies from current or historically important language designs:
  – data structures: pair/struct/record types; union/variant types; recursive types
  – data abstractions: generics; ad hoc vs. parametric polymorphism
  – control structures: goto, loops, case/switch, exceptions and exception handling
  – control abstractions: procedure call/return; function types; functions as data;
  – reactive programming; event dispatch; callbacks; stream programming
  – design dimensions: eager vs. lazy evaluation; purity vs. side-effects
  – objects, classes, modules, interfaces, mixins/traits
  – subtyping, inheritance, multiple inheritance
  – overloading, syntactic abstraction, macros
• Advanced topics/guest lectures on concurrent, parallel, distributed programming, security, or verification (linking to UG4 year courses)

Relevant QAA Computing Curriculum Sections: Comparative programming languages.

Transferrable skills: [No additional explicitly taught transferrable skills]

Reading List:
• Essentials of Programming Languages, Friedman, Wand and Haynes
• Practical foundations for programming languages, Harper

In addition online resources or readings specific to the programming language used for the course should be recommended, but need not appear in the course descriptor, for example:
• Programming F# 3.0. C. Smith, O'Reilly, 2013
• Functional Programming in Scala, P. Chisuano and R. Bjarnason, 2015

Breakdown of Learning and Teaching Activities: 2 lecture hours and 1 tutorial hour each week, with 3 coursework assignments.

<table>
<thead>
<tr>
<th>Activity</th>
<th>Hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lecture Hours</td>
<td>20</td>
</tr>
<tr>
<td>Seminar/Tutorial Hours</td>
<td>8</td>
</tr>
<tr>
<td>Supervise practical/Workshop/Studio hours</td>
<td>2</td>
</tr>
<tr>
<td>Summative assessment hours</td>
<td>25</td>
</tr>
<tr>
<td>Feedback/Feedforward hours</td>
<td>2</td>
</tr>
<tr>
<td>Directed Learning and Independent Learning hours</td>
<td>43</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
</tr>
</tbody>
</table>

Keywords: programming languages, interpreters, domain-specific languages, types, abstraction, parameterisation, exceptions, generics.
3 Course materials

[these should be prepared once the Board meeting at which the proposal will be discussed has been specified ]

3.1 Sample exam question
3.2 Sample coursework specification
3.3 Sample tutorial/lab sheet question
3.4 Any other relevant materials
4 Course management

[this information can be compiled in parallel to the elicitation of comments for section 5. ]

4.1 Course information and publicity

4.2 Feedback

4.3 Management of teaching delivery
5 Comments

5.1 Year Organiser Comments
5.2 Degree Programme Co-Ordinators
5.3 BoS Academic Secretary