

Chapter 1

Elpi introduction

In this chapter we will show how Elpi together with Coq-Elpi can be used to create new tactics. We will do this by giving a tutorial on how to implement the **iIntros** tactic from Iris.

We implement our tactic in the λ Prolog language Elpi [Dun+15; GCT19]. Elpi implements λ prolog [MN86; Mil+91; BBR99; MN12] and adds constraint handling rules to it [Mon11]. constraint handling will be explained in Section ?.

TODO: Defer constraint handling to later

To use Elpi as a Coq meta programming language, there exists the Elpi Coq connector, Coq-Elpi [Tas18]. We will use Coq-Elpi to implement the Elpi variant of **iIntros**, named above as **eiIntros**.

Our Elpi implementation **eiIntros** consists of three parts as seen in figure 1.1. The first two parts will interpret the DSL used to describe what we want to introduce. Then, the last part will apply the interpreted DSL. In section 1.1 we describe how a string is tokenized by the tokenizer. In section 1.2 we describe how a list of tokens is parsed into a list of intro patterns. In section 1.3 we describe how we use an intro pattern to introduce and eliminate the needed connectives. In every section we describe more parts of the Elpi programming language and the Coq-Elpi connector starting with the base concepts of the language and working up to the mayor concepts of Elpi and Coq-Elpi.

1.1 Tokenizer

The tokenizer takes as input a string. We will interpret every symbol in the string and produce a list of tokens from this string. Thus, the first step is to define our tokens. Next we show how to define a predicate that transform our string into the tokens we defined.

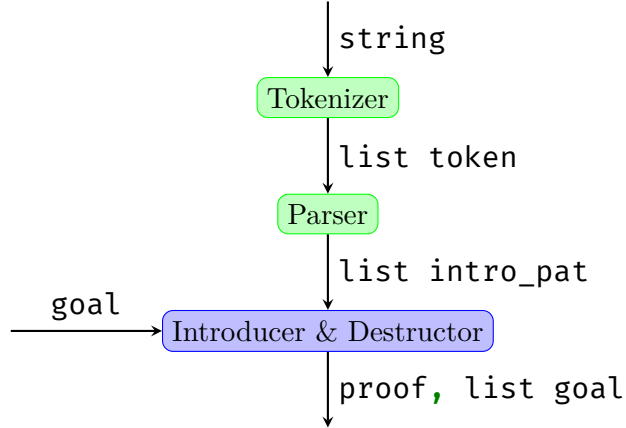


Figure 1.1: Structure of `eiIntros` with the input and output types on the edges.

1.1.1 Data types

We have separated the introduction patterns into several distinct tokens. Most tokens just represent one or two characters, but some tokens also contain some data associated with that token. For example `"H1"` is tokenized as the name token containing the string `"H1"`.

```

1  kind token type.
2
3  type tAnon, tFrame, tBar, tBracketL, tBracketR, tAmp,
4      tParenL, tParenR, tBraceL, tBraceR, tSimpl,
5      tDone, tForall, tAll token.
6  type tName string -> token.
7  type tNat int -> token.
8  type tPure option string -> token.
9  type tArrow direction -> token.
10
11 kind direction type.
12 type left, right direction.

```

We first define a new type called `token` using the `kind` keyword, where `type` specifies the kind of our new type. Then we define several constructors for the `token` type. These constructors are defined using the `type` keyword, we specify a list of names for the constructors and then the type of those constructors. The first set of constructors do not take any arguments, thus have type `token`, and just represent one or more constant characters. The next few constructors take an argument and produce a to-

ken, thus allowing us to store data in the tokens. For example, `tName` has type `string -> token`, thus containing a string. Besides `string`, there are a few more basic types in Elpi such as `int`, `float` and `bool`. We also have higher order types, like `option A`, and later on `list A`.

```
1 kind option type -> type.
2 type none option A.
3 type some A -> option A.
```

Creating types of kind `type -> type` can be done using the `kind` directive and passing in a more complicated kind.

We can now represent a string as a list of these tokens. Given the string `"[H %H']"` we can represent it as the following list of type `token`:

```
1 [tBracketL, tName "H", tPure (some "H'"), tBracketR]
```

1.1.2 Predicates

Programs in Elpi consist of predicates. Every predicate can have several rules to describe the relation between its inputs and outputs.

```
1 pred tokenize i:string, o:list token.
2 tokenize S O :-
3   rex.split "" S SS,
4   tokenize.rec SS O.
```

Line 1 describes the type of the predicate. The keyword `pred` starts the definition of a predicate. Next we give the name of the predicate, "tokenize". Lastly, we give a list of arguments of our predicate. Each argument is marked as either `i:`, they act as an input or `o:`, they act as an output, in section 1.1.3 a more precise definition is given. In the only rule of our predicate, defined on line 2, we assign a variable to both of the arguments. `S` has type `string` and is bound to the first argument. `O` has type `list token` and is bound to the second argument. By calling predicates after the `:-` symbol we can define the relation between the arguments. The first predicate we call, `rex.split`, has the following type:

```
1 pred rex.split i:string, i:string, o:list string.
```

When we call it, we assign the empty string to its first argument, the string we want to tokenize to the second argument, and we store the output list of string in the new variable `SS`. This predicate allows us to split a string at a certain delimiter. We take as delimiter the empty string, thus splitting the string up in a list of strings of one character each. Strings in Elpi are

based on OCaml strings and are not lists of characters. Since Elpi does not support pattern matching on partial strings, we need this workaround.

The next line, line 4, calls the recursive tokenizer, `tokenizer.rec`¹, on the list of split string and assigns the output to the output variable `O`.

The reason predicates in Elpi are called predicates and not functions, is that they don't always have to take an input and give an output. They can sometimes better be seen as predicates defining for which values of their arguments they hold. Each rule defines a list of predicates that need to hold for their premise to hold. Thus, a predicate can have multiple values for its output, as long as they hold for all contained rules. These multiple possible values can be reached by backtracking, which we will discuss in section 1.1.5. To execute a predicate, we thus find the first rule for which its premise is sufficient for the arguments we supply. We then check if each of the predicates in the conclusion hold starting at the top. If they hold, and we get a value for every output argument, we are done executing our predicate. How we determine when arguments are sufficient and what happens when a rule does not hold, we will discuss in the next two sections.

1.1.3 Matching and unification

The arguments of a predicate can be more than just a variable. We can supply a value containing variables and depending on the argument mode, input or output, we match or unify the input with the premise respectively.

`tokenize.rec` uses matching and unification to solve most cases.

```

1  pred tokenize.rec i:list string, o:list token.
2  tokenize.rec [] [] :- !.
3  tokenize.rec [" " | SL] TS :- !, tokenize.rec SL TS.
4  tokenize.rec ["$" | SL] [tFrame | TS] :- !,
5      tokenize.rec SL TS.
6  tokenize.rec ["/", "/", "=" | SL] [tSimpl, tDone | TS] :- !,
7      tokenize.rec SL TS.
8  tokenize.rec ["/", "/" | SL] [tDone | TS] :- !,
9      tokenize.rec SL TS.
```

This predicate has several rules, we chose a few to highlight here. The first rule, on line 2, has a premise and a cut as its conclusion, we will discuss cuts in section 1.1.5, for now they can be ignored. This rule can be used when the first argument matches `[]` and if the second argument unifies with `[]`. The difference is that, for two values to match they must have the exact same constructors and can only contain variables in the same places in the

¹Names in Elpi can have special characters in them like `.`, `-` and `>`, thus, `tokenize` and `tokenize.rec` are fully separate predicates. It is just a convention that when creating a helper predicate we name it by adding a dot and a short name for the helper.

value. Thus, the only valid value for the first argument of the first rule is `[]`. When unifying two values we allow a variable to be unified with a constructor, when this happens the variable will get assigned the value of the constructor. Thus, we can either pass `[]` to the second argument, or some variable `V`. After the execution of the rule the variable `V` will have the value `[]`.

The next four rules use the same principle. They use the list pattern `[E1, ..., En | TL]`, where `E1` to `En` are the first n values and `TL` is the rest of the list, to match on the first few elements of the list. We unify the output with a list starting with the token that corresponds to the string we match on. The tails of the input and output we pass to the recursive call of the predicate to solve.

When we encounter multiple rules that all match the arguments of a rule we try the first one first. The rules on line 6 and 8 would both match the value `["/", "/", "="]` as first argument. But, we interpret this use the rule on line 6 since it is before the rule on line 8. This results in our list of strings being tokenized as `[tSimpl, tDone]`.

A fun side effect of output being just variables we pass to a predicate is that we can also easily create a function that is reversible. If we change the mode of our first argument to output and move rule 3 to the bottom, we can pass in a list of tokens and get back a list of strings representing this list of tokens.

Question: Don't know what to do with this, but is an interesting fact and shows the versatility, we might use it later.

1.1.4 Functional programming in Elpi

While our language is based on predicates we still often defer to a functional style of programming. The first language feature that is very useful for this goal is spilling. Spilling allows us to write the entry point of the tokenizer as defined in section 1.1.2 without the need of the temporary variable to pass the list of strings around.

```
1 pred tokenize i:string, o:list token.
2 tokenize S 0 :- tokenize.rec {rex.split "" S} 0.
```

We spill the output of a predicate into the input of another predicate by using the `{ }` syntax. We don't specify the last argument of the predicate and only the last argument of a predicate can be spilled. It is mostly equal to the previous version, but just written shorter. There is one caveat but it will be discussed in ?.

TODO: Refer to relevant section

The second useful feature is how lambda expressions are first class citizens of the language. A `pred` statement is a wrapper around a constructor definition using the keyword `type`, where all arguments are in output mode. The following predicate is equal to the type definition below it.

```

1  pred tokenize i:string, o:list token.
2  type tokenize string -> list token -> prop.

```

The **prop** type is the type of propositions, and with arguments they become predicates. We are thus able to write predicates that accept other predicates as arguments.

```

1  pred map i:list A, i:(A -> B -> prop), o:list B.
2  map [] _ [].
3  map [X|XS] F [Y|YS] :- F X Y, map XS F YS.

```

map takes as its second argument a predicate on **A** and **B**. On line 3 we map this predicate to the variable **F**, and we then use it to either find a **Y** such that **F X Y** holds, or check if for a given **Y**, **F X Y** holds. We can use the same strategy to implement many of the common functional programming higher order functions.

1.1.5 Backtracking

In this section we will finally describe what happens when a rule fails to complete halfway through. We start with a predicate which will be of much use for the last part of our tokenizer.

```

1  pred take-while-split i:list A, i:(A -> prop),
2  o:list A, o:list A.
3  take-while-split [X|XS] Pred [X|YS] ZS :- Pred X,
4  take-while-split XS Pred YS ZS.
5  take-while-split XS _ [] XS.

```

take-while-split is a predicate that should take elements of its input list till its input predicate no longer holds and then output the first part of input in its third argument and the last part of the input in its fourth argument.

The predicate contains two rules. The first rule, defined on lines 2 and 3, recurses as long as the input predicate, **Pred** holds for the input list, **[X|XS]**. The second rule returns the last part of the list as soon as **Pred** no longer holds.

The first rule destructs the input in its head **X** and its tail **XS**. It then checks if **Pred** holds for **X**, if it does, we continue the rule and call **take-while-split** on the tail while assigning **X** as the first element of the first output list and the output of the recursive call as the tail of the first output and the second output. However, if **Pred X** does not succeed we backtrack to the previous rule in our conclusion. Since there is no previous rule in the conclusion we instead undo any unification that has happened

and try the next possible rule. This will be the rule on line 4 and returns the input as the second output of the predicate.

We can use `take-while-split` to define the rule for the token `tName`

```

1 type tName string -> token.
2
3 tokenize.rec SL [tName S | TS] :-
4   take-while-split SL is-identifier S' SL',
5   { std.length S' } > 0, !,
6   std.string.concat "" S' S,
7   tokenize.rec SL' TS.

```

To tokenize a name we first call `take-while-split` with as predicate `is-identifier`, which checks if a string is valid identifier character, whether it is either a letter or one of a few symbols allowed in identifiers. It thus splits up the input string list into an list of string that is a valid identifier and the rest of the input. On line 5 we check if the length of the identifier is larger than 0. We do this by spilling the length of `S'` into the `>` predicate. Next, on line 6, we concatenate the list of strings into one string, which will be our name. And on line 7, we call the tokenizer on the rest of the input, to create the rest of our tokens.

If our length check does not succeed we backtrack to next rule that matches, which is

```

1 tokenize.rec XS _ :- !,
2   coq.say "unrecognized tokens" XS, fail.

```

It prints an error messages saying that the input was not recognized as a valid token, after which it fails. The predicate thus does not succeed. There is one problem, if line 6 or 7 fails for some reason in the `tName` rule of the tokenizer, the current input starting at `X` is not unrecognized as we managed to find a token for the name at the start of the input. Thus, we don't want to backtrack to another rule of `tokenize.rec` when we have found a valid name token. This is where the cut symbol, `!`, comes in. It cuts the backtracking and makes certain that if we fail beyond that point we don't backtrack in this predicate.

If we take the following example

```

1 tokenize.rec ["H", "^"] TS
2           ↓ calls
3 tokenize.rec ["^"] TS'

```

When evaluating this predicate we would first apply the name rule of the `tokenize.rec` predicate. This would unify `TS` with `[tName "H" | TS']`

and call line 3, `tokenize.rec ["^"] TS'`. Every rule of `tokenize.rec` fails including the last fail rule. This rule does first print `"unrecognized tokens ^"` but then also fails. Now when executing the rule of line 1, we have failed on the last predicate of the rule. If there was no cut before it, we would back-track to the fail rule and also print `"unrecognized tokens [H, ^]"`. But, because there is a cut we don't print the faulty error message. Thus, we only print meaningful error message when we fail to tokenize an input.

1.2 Parser

1.2.1 Data structure

Explain the data structure of intro patterns

- Tree structure
- iList, allow for destructing of more than just or, and.

```

1 kind ident type.
2 type iNamed string -> ident.
3 type iAnon term -> ident.
4
5 kind intro_pat type.
6 type iFresh, iSimpl, iDone intro_pat.
7 type iIdent ident -> intro_pat.
8 type iList list (list intro_pat) -> intro_pat.
```

1.2.2 Reductive descent parsing

- We can translate a grammar directly to a parser
- Give partial grammar for the intro patterns (in terms of tokens?)
- Show code for parser, use `parse_elist` as example
- How to translate a grammar rule to a parser rule

$$\begin{aligned} \langle \text{intro_pattern_list} \rangle &::= \epsilon \\ &\quad | \quad \langle \text{intro_pattern} \rangle \langle \text{intro_pattern_list} \rangle \end{aligned}$$

$$\begin{aligned} \langle \text{intro_pattern} \rangle &::= \langle \text{ident} \rangle \\ &\quad | \quad '?' \mid '\neq' \mid '//' \\ &\quad | \quad '[' \langle \text{intro_pattern_list} \rangle ']' \\ &\quad | \quad '(' \langle \text{intro_pattern_conj_list} \rangle ')' \end{aligned}$$

$$\begin{aligned} \langle \text{intropattern_list} \rangle &::= \epsilon \\ &| \langle \text{intropattern} \rangle ' | ' \langle \text{intropattern_list} \rangle \\ &| \langle \text{intropattern} \rangle \langle \text{intropattern_list} \rangle \end{aligned}$$

$$\begin{aligned} \langle \text{intropattern_conj_list} \rangle &::= \epsilon \\ &| \langle \text{intropattern} \rangle ' \& ' \langle \text{intropattern_conj_list} \rangle \end{aligned}$$

```

1  pred parse_ip i:list token, o:list token, o:intro_pat.
2  parse_ip [tAnon | TS] TS (iFresh) :- !.
3  parse_ip [tSimpl | TS] TS (iSimpl) :- !.
4  parse_ip [tDone | TS] TS (iDone) :- !.
5  parse_ip [tName X | TS] TS (iIdent (iNamed X)) :- !.

```

```

1  parse_ip [tBracketL | TS] TS' (iList L) :- !,
2  parse_ilst TS [tBracketR | TS'] L.
3  parse_ip [tParenL | TS] TS' IP :- !,
4  parse_conj_ilst TS [tParenR | TS'] L',
5  {std.length L'} ≥ 2,
6  foldr {std.drop-last 2 L'} (iList [{std.take-last 2 L'}]) (x\ a\ r\ r =

```

```

1  pred parse_ilst i:list token, o:list token, o:list (list intro_pat).
2  parse_ilst [tBracketR | TS] [tBracketR | TS] [[]].
3  parse_ilst TS R [[IP] | LL'] :-
4    parse_ip TS [tBar | RT] IP,
5    parse_ilst RT R LL'.
6  parse_ilst TS R [[IP | L] | LL'] :-
7    parse_ip TS RT IP,
8    parse_ilst RT R [L | LL'].
9
10 pred parse_conj_ilst i:list token, o:list token, o:list intro_pat.
11 parse_conj_ilst TS [tParenR | R] [IP] :-
12   parse_ip TS [tParenR | R] IP.
13 parse_conj_ilst TS R [IP | L'] :-
14   parse_ip TS [tAmp | RT] IP,
15   parse_conj_ilst RT R L'.

```

1.2.3 Danger of backtracking

Explain how backtracking can lead to massive slowdowns when not managed correctly.

- Show timing of current `parse_elist` code on larger inputs
- Change backtracking
- Show new timings

```

1  pred parse_elist i:list token, o:list token, o:list (list intro_pat).
2  parse_elist [tBracketR | TS] [tBracketR | TS] [[]].
3  parse_elist TS R [IPS | LL'] :-
4    parse_ip TS RT IP,
5    (
6      (
7        RT = [tBar | RT'],
8        parse_elist RT' R LL',
9        IPS = [IP]
10     );
11    (
12      parse_elist RT R [L | LL'],
13      IPS = [IP | L]
14    )
15  ).

```

1.3 Applier

Something about Coq-Elpi becoming relevant here

1.3.1 Elpi coq HOAS

1+1

```

1  app [global (const «Nat.add»),
2      app [global (indc «S»), global (indc «0»)],
3      app [global (indc «S»), global (indc «0»)]]

forall (n: nat), n + 1

1  prod `n` (global (indt «nat»)) c0 \
2    app [global (indt «sum»),
3        c0,
4        app [global (indc «S»), global (indc «0»)]]

```

1.3.2 Quotation and anti-quotation

```
1 {{ forall (n: nat), n + 1 }} =
2   prod `n` (global (indt «nat»)) c0 \
3     app [global (indt «sum»),
4         c0,
5         app [global (indc «S»), global (indc «0»)]]
```

```
1 Plus = {{ 2 }},
2 Foo = {{ forall (n: nat), n + lp:Plus }}
```

1.3.3 Coq context in elpi

1.3.4 Proofs in Elpi

```
1 pred do-iStartProof i:hole, o:ihole.
2 do-iStartProof (hole {{ let _ := _ in _ }} _) _ :- !,
3   coq.error "iStartProof: goal is a `let`, use `simpl`, `intros x`, `iI
4 do-iStartProof (hole {{ @envs_entails lp:PROP (@Envns lp:PROPE lp:CI lp:
5 do-iStartProof (hole Type Proof) (ihole N (hole NType NProof)) :-
6   coq.elaborate-skeleton {{ as_emp_valid_2 lp:Type _ (tac_start _ _) }}
7   Proof = {{ as_emp_valid_2 _ _ (tac_start _ lp:NProof) }},
8   coq.typecheck NProof NType ok,
9   NType = {{ envs_entails (Envns _ _ lp:N) _ }}.
```

1.3.5 Introducing anonymous Iris hypothesis

```
1 pred do-iIntros i:(list intro_pat), i:ihole, i:(ihole -> prop).
2 do-iIntros [] IH C :- !, if-debug (coq.say "do-iIntros done"), C IH.
3 do-iIntros [iFresh | IPS] (ihole N H) C :- !,
4   increase-ctx-count N NS,
5   do-iIntro-ident (iAnon N) (ihole NS H) IH, !,
6   do-iIntros IPS IH C.
```

1.3.6 Continuation Passing Style

```
1 pred do-intro-anon i:hole, i:(hole -> prop).
2 do-intro-anon (hole Type Proof) C :-
```

```

3   coq.ltac.fresh-id "a" {{ False }} ID,
4   coq.id->name ID N,
5   coq.elaborate-skeleton (fun N _ _> Type Proof ok,
6   Proof = (fun _ T IntroFProof),
7   @pi-decl N T x\
8     coq.typecheck (IntroFProof x) (F x) ok,
9     C (hole (F x) (IntroFProof x))).

```

1.3.7 Backtracking in proofs

```

1   pred do-iIntro-ident i:ident, i:ihole, o:ihole.
2   do-iIntro-ident ID (ihole N (hole Type Proof)) (ihole N (hole IType Int
3     ident->term ID _ T,
4     coq.elaborate-skeleton {{ tac_impl_intro _ lp:T _ _ _ _ }} Type
5     Proof = {{ tac_impl_intro _ _ _ _ _ _ lp:IProof }}},
6     coq.typecheck IProof IType' ok,
7     pm-reduce IType' IType,
8     if (IType = {{ False }}) (coq.error "eiIntro: " X " not fresh") (true
9   do-iIntro-ident ID (ihole N (hole Type Proof)) (ihole N (hole IType Int
10    ident->term ID _ T,
11    coq.elaborate-skeleton {{ tac_wand_intro _ lp:T _ _ _ _ }} Type Pro
12    Proof = {{ tac_wand_intro _ _ _ _ _ _ lp:IProof }}},
13    coq.typecheck IProof IType' ok,
14    pm-reduce IType' IType,
15    if (IType = {{ False }}) (coq.error "eiIntro: " X " not fresh") (true
16  do-iIntro-ident ID _ _ :-
17    ident->term ID X _,
18    coq.error "eiIntro: " X " could not introduce".

```