Fiat Lingua

Title: Designing an Artificial Language: Anaphora

Author: Rick Morneau

MS Date: 07-12-1994

FL Date: 07-01-2019

FL Number: FL-00005E-00

Citation: Morneau, Rick. 1994. "Designing an Artificial Language: Anaphora." FL-00005E-00, *Fiat Lingua*, <http://fiatlingua.org>. Web. 01 July 2019.

Designing an Artificial Language:

Anaphora

by Rick Morneau

October, 1993
Revised July 12, 1994

Copyright © 1991, 1994 by Richard A. Morneau,
all rights reserved.

```
[The following essay is a compilation of several items that
I posted to the Conlang mailing list in October 1993. The
Conlang mailing list is dedicated to the discussion of the
construction of artificial languages. To subscribe, send an
email message with the single line:

            SUBSCRIBE CONLANG your name
  to LISTSERV@BROWNVM.BROWN.EDU. I would like to thank And
  Rosta for starting the discussion. I would also like to
  thank And Rosta, Jacques Guy, Colin Fine, and Prentiss
            Riddle for many valuable comments.]
```

The design of a comprehensive yet simple anaphoric system is not especially difficult. All natural languages have one. However, as is typical of natural languages, the anaphoric systems are clouded in idiosyncrasy and irregularity.

One of the problems that many people have is that they tend to think of anaphora as belonging to a special, closed class of words. In English, we think of third person pronouns ("he", "she", "it", etc.), demonstratives ("this", "those", etc.), auxiliaries ("be", "have", and "do", etc.) and a handful of oddballs ("herself", "each other", "so", "such", etc.) as most of the available anaphora. Here are some examples:

```
I love anchovy ice cream. Do you? (Anaphor: "do")

William Shakespeare lived in a small town with his pet rock
and his wife Fifi Yokohama. He would not eat veggies, she
would not eat vegemite, and IT didn't eat at all.
(Anaphora: "his", "he", "she" and "IT")
```

```
John said he'll definitely attend the class on Creative
Suffering. Louise will too. (Anaphor: "will")
```

However, these "closed class anaphora" are not the only ones. Consider the following:

```
1. Ten theoretical physicists and eight sanitary engineers
attended the seminar. They were constantly heckling them.
```

Obviously, we can't use the anaphora "they" and "them" in the second sentence of (1). Instead, we need something like:

```
2. The engineers were constantly heckling the physicists.
```

The point, though, is that the words "engineers" and "physicists" in (2) are anaphora, and they can continue to be used as such throughout the remainder of the dialog. Thus, the head word of a phrase is used as a referent for the entire phrase. I'll call these "open class anaphora".

[For the GB nitpickers: Obviously, I am using the word "anaphor" in a loose functional sense, rather than in a strict syntactic sense. Whether an anaphor is legal because node A c-commands node B is not really relevant to this discussion.]

Sometimes, especially when writing, we define new open class anaphora explicitly, as in:

```
3. This contract is between Timothy TackyTie (henceforth
the first party) and Wendall WeeWilly (henceforth the
second party)…
```

In (3), the anaphora are explicitly defined as "the first party" and "the second party". But we can also do it in informal writing and speech:

```
4. Ten computational linguists and twelve theoretical
linguists attended the seminar. The comps were constantly
heckling the theos. Finally, the theos got so angry that
they mooned the comps and left.
```

Another common way to create open class anaphora is to use single letters or abbreviations:

```
5. In discussing the "Best Artificial Language Linguists
Ever Designed" (BALLED), the designers forgot that there
were many other lingwackos out there, who were out to get
BALLED and who would ridicule it at every opportunity.
```

Of course, once an abbreviation becomes recognizable without introduction, it will no longer be an anaphor - it will be a proper noun (like USA, IBM, etc.).

The major difference between the open (O) and closed (C) classes of anaphora is that the Os tend to keep their referents throughout the discourse, while the referents of the Cs are constantly changing. Thus, the anaphor "BALLED" in (5) will refer to the same thing throughout the dialog, while anaphora such as "he", "do" or "each other" will continually take on new meanings.

One other thing should be mentioned. Most anaphora are "backward-referring"; that is, the anaphor refers to something that was mentioned earlier. It is also possible to have "forward-referring" anaphora, as in:

> 6. After ordering a pint of his favorite ale, Robert was
> perplexed when the barmaid replied that the fishmonger was
> next door. The Great English Vowel Shift had begun.

In (6), "his" precedes its referent "Robert".

So, how do you handle anaphora in an artificial language (henceforth AL)? One solution would be to create a lot of noun and verb classes. But this will not always solve the problem. You'll often have situations where you want to differentiate between two or more members of the same class (such as "physicists" and "engineers").

A better way, in my opinion, is to design the phonotactics and morphotactics of your AL to allow the head word of any phrase to be contracted or to be combined in some way with an important modifier. The result would always be immediately recognizable as an anaphor by its form. The contraction could then be used as an anaphor for the entire phrase from that point on. (You could modify this rule to allow the contraction to take on a new meaning if its pattern matches a newly introduced phrase.) Here's how something like this might sound in English:

> The Sheboygan Bandits and the Milwaukee Dragoons faced off
> at Lovemud Stadium on Sunday. The Mil'goons beat the
> She'its out of their expected title.

Unfortunately, English is not really suited for this. An AL, however, can be designed to allow such an anaphoric system, and there are many ways to do it. Here's one possible approach... Let open class words have the following form:

> stem + classifier + part-of-speech

where

```
stem = [CV]
classifier = CC
part-of-speech = V

[] means 1 or more of the enclosed item
C is any consonant
V is any vowel ('o'=noun, 'e'=verb, etc.)
```

Thus, examples of open class words would be **mande, kitusta, jonabefti**, etc. (You would probably want to exercise some restraint in your choice of legal consonant clusters to make pronunciation as easy as possible.)

There should be two types of anaphor: simple anaphora and compound anaphora. A simple anaphor will be formed from one or more initial syllables of the head word followed by a glottal stop (represented here by an apostrophe) followed by the final part-of-speech vowel. Thus, simple anaphora will have the structure:

```
[CV]'V
```

where both parts of the anaphor are taken from the head word.

Compound anaphora will be formed from one or more initial syllables of a significant **modifier** of the head word, followed by a glottal stop followed by the final VCCV of the head word. Thus, compound anaphora will have the structure:

```
[CV]'VCCV
```

where the first part of the anaphor comes from a significant modifier of the head word, and the second part of the anaphor comes from the head word itself. (Note that both forms are consistent with a self-segregating morphology.)

For example, consider the following sample noun phrase:

```
timanodendo janasuski tupya
engineer    sanitary   ten
"ten sanitary engineers"
```

The above example could have the simple anaphora *ti'o, tima'o* or *timano'o*, or the compound anaphora *ja'endo, jana'endo* or *janasu'endo*.

And the following verb phrase:

4

```
jujushimpe makitundo  bubuski
heckle     fishmonger illiterate
"to heckle illiterate fishmongers"
```

could be abbreviated to the simple anaphora *ju'e, juju'e* or *jujushi'e*, or to the the compound anaphora *ma'impe, maki'impe* or *makitu'impe*.

During the discussion that took place on the conlang list, And Rosta took me to task because my proposed anaphoric system could not deal with the following kind of problem:

> ```
> A dog was attracted to a dog. But its owner kept it away
> from it.
> ```

I agree that the proposed system cannot deal with this kind of situation, but I don't understand why anyone would **want** an anaphoric system to be able to deal with it. This kind of situation will only be used when the speaker is being humorous or intentionally ambiguous. As far as I'm concerned, if the speaker wants to have fun, then let him! Besides, you could always distinguish between "the first dog" and "the second dog", or "the former" and "the latter". In my opinion, this is a non-problem, and I see no reason to waste time on it.

However, we most certainly **can** deal with a more reasonable version of this sentence, such as:

> ```
> A big dog was attracted to a little dog. But its owner kept
> it away from it.
> ```

Using compound anaphora, one possible permutation would be:

> ```
> A big dog was attracted to a little dog. But li'og's owner
> kept bi'og away from li'og.
> ```

One other problem that cropped up in the discussion had to do with resolving the individual referents of a phrase that implicitly referred to more than one referent. For example, does the phrase "two identical twins", provide a single referent or a double referent? How about the phrase "box of nuts and bolts" or "ten million civilians"?

I strongly feel that a properly designed anaphoric system should be able to provide an unambiguous index to any referent. The system I proposed does this very well. Furthermore, if the referent is ambiguous, then the anaphor should also be ambiguous. In other words, the anaphoric system should not be given the additional duty of disambiguating an ambiguous phrase. Disambiguation should be handled explicitly by the speaker. Thus, "a dog and a dog" is

intentionally ambiguous (in addition to being unnatural). I do not feel that an anaphoric system should be required to resolve an intentional ambiguity.

In the case of "two identical twins", only one referent was provided, and the system proposed here can deal with it very well. The referent is "two identical twins", and one possible anaphor would be "id'ins".

Now, some people feel that the anaphoric system must also provide an unambiguous index to **each** of the twins. If so, then the anaphoric system must provide an index to a referent that has not even been mentioned. If neither of the twins has been mentioned separately, then the referent does not exist, and I see no reason to provide an index to a non-existent referent.

In other words, what some people seem to want is an anaphoric system that can also provide *semantic decomposition*. I do not feel that this should be the purpose of an anaphoric system, even though it is occasionally possible in natural languages. Considering the many, many possible kinds of groupings (twins, clubs, choirs, companies, orchards, boxes of spare parts, etc), such a system would be very complex, and I'm not even sure if it would be possible.

In summary, I feel that an anaphoric system should be rich enough to provide an unambiguous index to any unambiguous referent. Such a system should **not** have the additional duties of disambiguation or semantic decomposition.

Finally, keep in mind that the approach to designing anaphora discussed here will work best if the phonotactics and morphotactics of your AL are designed for it. Unfortunately, if you are designing a yachecle (Yet Another CHauvinistic Euro-CLonE), the above system may not be very practical. :-(

<center>End of Essay</center>

[Postscript: I would like to emphasize that the above essay reflects **my** opinions on how to deal with anaphora in the design of ALs, and that others who took part in the discussion do not necessarily agree with me. For example, some people felt that an anaphoric system **should** be designed to provide disambiguation and semantic decomposition.]