

# THE EMERGENCE OF THE INSUFFICIENCY IF

## GUTENBERG AUF! IN ACTION: A COMPUTATIONAL, EMPIRICAL APPROACH TO THE STUDY OF LANGUAGE EVOLUTION

*Research in Progress*

J. D. Alexandre

R. E. Jennings

Laboratory for Logic and Experimental Philosophy

Simon Fraser University

British Columbia, Canada

ABSTRACT. We examine a dualized form of the English connective *if*, and give an account of its potential origins and evolution, within the framework of Jennings' mutation hypothesis<sup>1</sup>. In Jennings' language model, meanings are physical types, and thus species: unions of populations that are temporally ordered along an engendering relation. As this engendering relationship between generations is not strictly chronological, but rather statistical, firm conclusions cannot be drawn on the basis of a cursory examination of a small subset of the extant historical language data. The method we adopt here arose out of a perceived need for a computational aid in the production of solid, empirical results in the area of the biology and evolution of language, in particular the genealogy of connective vocabulary in natural languages. We have been developing software, *Gutenberg Auf!*, to assist in the analysis of a large corpus of historical English literature, which we have taken from two complementary sources: Project Gutenberg<sup>2</sup>, and Early English Books Online<sup>3</sup>. In this paper we report the first application of the software: the confirmation of an earlier<sup>4</sup> propounded hypothesis about the emergence in English of the "insufficiency *if*" (the *if* of *even if*).

---

<sup>1</sup> See R.E. Jennings and Nathalie Schapansky. "Without: from separation to negation; A Case Study in Logicalization". *Proceedings of the 2000 Annual Conference of the Canadian Linguistics Association*. pp 147-157.

<sup>2</sup> Project Gutenberg (<http://www.gutenberg.net>), a volunteer organization committed to the creation and distribution of public domain literary e-texts, contains a massive body of texts, primarily in English, spanning the past 800 years. This extensive collection of machine-readable text is a goldmine for historical linguistic research. We have downloaded approximately 7000 texts (3 gigabytes), and indexed them by title, author name, author lifespan, publication date, publication location, and other variables relevant to the automated search and analysis process. Works from the Gutenberg corpus we cite by filename, for example (GC: zadig10a.txt).

<sup>3</sup> Early English Books Online (<http://eebo.chadwyck.com>) is a collection of approximately 5000 full text English books published prior to 1700, along with scanned images of the manuscript pages. It serves as a useful counterpart to the Gutenberg Corpus, the bulk of which comes from the 18<sup>th</sup> and 19<sup>th</sup> centuries.

<sup>4</sup> See R.E. Jennings. "Semantics and Explanation" in . . . .



## 1. Genesis

In the case of conditionals, as opposed to sentences formed with many of the other connectives, there are no clearly established conventions to specify the scope of negations, as there is no clean, unambiguous way to negate an entire conditional (without resorting to the philosopher's unwieldy "it is not the case that..."). Thus, the *not* tends to be applied to the main verb, in the consequent clause. The insufficiency *if*, we propose, originated in a hearer's misconstrual of an internally negated conditional, e.g.

If you stand out there all night, I won't let you in.

(with a reading, presumably assisted by prosody, of *It's not the case that if you stand out there all night I will let you in*) as an un-negated conditional in which only the consequent has been negated. That is, upon hearing a conditional of the intended form

Not(A *if<sub>s</sub>* B)

the scope of the negation is construed as including only the antecedent:

Not(A) *if<sub>i</sub>* B

In order for the utterance as a whole, under this scope-foreshortened reading, to maintain the same occasions of use as the original, and thus evade correction, the hearer must establish a new meaning for the connective *if*, one that we denote here as *if<sub>i</sub>*, the insufficiency *if*, in contrast to the historically primary sufficiency *if* meaning, which we denote *if<sub>s</sub>*.

## 2. Incubation

During this stage, the new construal remains hidden beneath the old, and the novel meaning is reinforced by its successful production of correct occasions of use for the vocabulary. In the case of *if*, speakers who had misconstrued *Not(A if<sub>s</sub> B)* as *Not(A) if<sub>i</sub> B* go on to use *if<sub>i</sub>* in novel utterances successfully, but still in sentential environments in which an *if<sub>s</sub>* reading is also possible.

## 3. Migration

Once a sufficient number of speakers have adopted the *if<sub>i</sub>* reading for the *if* of negated conditionals, the linguistic climate becomes more forgiving and allows *if<sub>i</sub>* to migrate to contexts in which no negation is present. Early migrations may at first be more likely to succeed in contexts where a negation is still implicitly present, perhaps absorbed into the meaning of the main verb or predicate, as in cases such as

If she had wanted to say something, she kept silent.  
(that is, *despite the possibility that she may have wanted to say something, she did not*)

where keeping silent contains an implicit negation of her having said something. Eventually, *if* begins to occur in environments where it *requires* an *if<sub>i</sub>* meaning; that is, it occurs in utterances for which there is no plausible reading under which the *if* retains an *if<sub>s</sub>* meaning.

#### 4. Ambiguity

Once *if<sub>i</sub>* has firmly established itself in the linguistic community, speakers and listeners will come up against cases in which the *if* could be plausibly given either an *if<sub>s</sub>* or an *if<sub>i</sub>* reading, leading to an ambiguity between cases such as

If<sub>s</sub> he wins the lottery, he'll be MISERABLE.  
(Winning will make him miserable)

If<sub>i</sub> he wins the LOTTERY, he'll be miserable.  
(Winning won't cure his misery)

which cannot be distinguished in the absence of additional contextual or prosodic clues.

#### 5. Marking

In the face of ambiguity between *if<sub>s</sub>* and *if<sub>i</sub>*, we begin to see various types of marking appear to make explicit the *if<sub>i</sub>* reading. One of the earlier methods of marking appears to have been to introduce the consequent clause with *yet*, as in Shakespeare:

Or, if you cannot weep, yet give some groans,  
Though not for me, yet for your aching bones.<sup>7</sup>

This use of *yet* closely parallels our current use of *still*, as in

If the guilt was mine, I have suffered for it. If it was not mine, still I have suffered for it.<sup>8</sup>

The method of *if<sub>i</sub>*-marking that has become predominant in modern English, however, is the extremal *even* of *even if*, which we will examine in some detail. An example from Hume:

I am sensible, that nothing can be more unphilosophical than to be positive or dogmatical on any subject; and that, even if excessive scepticism could be maintained, it would not be more destructive to all just reasoning and inquiry.<sup>9</sup>

We would expect the historical introduction of this marked *if<sub>i</sub>* to have had a gradual, steady increase, paralleling, though offset by some period, the emergence of *if<sub>i</sub>* itself. We turn now to our first trial of *Gutenberg Auf!*, for empirical confirmation of this prediction.

In the Early English Books Online (EEBO) database, of the 910 full-text works it contains that were published prior to 1600, not a single instance of *even if<sup>10</sup>* is to be found, though not on account of any paucity of *if<sub>s</sub>*'s, of which, during this time period, EEBO contains over 100,000 instances. This absence of *even if* prior to the 17<sup>th</sup> century is corroborated by the Gutenberg corpus, in which not one of the 14,332 occurrences of *if* prior to 1600 was marked

---

<sup>7</sup> William Shakespeare (1564-1616), *History of Troilus and Cressida*, pub. 1602

<sup>8</sup> William Dean Howells (1837-1920), *A Foregone Conclusion* (GC: 8frcn10.txt)

<sup>9</sup> David Hume (1711-1776), *The Principles of Morals* (GC: nqpmr10.txt)

<sup>10</sup> This includes instances involving the archaic spelling *euen if*.

by *even*. Actually, this is not quite accurate; there was one text in the Gutenberg corpus, dated 1594, in which the following sentence occurred:

And again, over and besides the weakness and ill-furnishing of the rest,  
they were all so deeply laden, that they had not been able (*even if* they had been charged) to  
have held out any long fight. (italics added) <sup>11</sup>

Deciding to investigate, we turned to the manuscript images of a 1599 edition of the same work in the EEBO database, which read thus:

**And againe ouer and besides the weakenesse and the ill furnishing of the rest,  
they were all so deeply laden, that they had not bene able (if they had bene charged)  
to haue held out any long fight.**

And againe ouer and besides the weakenesse and the ill furnishing of the rest,  
they were all so deeply laden, that they had not bene able (if they had bene charged)  
to haue held out any long fight. <sup>12</sup>

One can only presume that a later editor, in the process of modernizing the spelling, took the liberty of inserting the *even*, most likely for the sake of clarity for readers who were by that time accustomed to this type of extremal marking of *if*.

During the 17<sup>th</sup> century, extremal *if* clauses began their gradual ascent, and very gradual it was. In the EEBO database, only 19 of some 600,000 occurrences of *if* are marked by the extremal *even* between 1600 and 1700. <sup>13</sup> The Gutenberg corpus, during these years, produces but one occurrence of *even if*, <sup>14</sup> and this from amongst 21,121 solitary *if*'s.

During the 18<sup>th</sup> century, however, the extremal *even if* seemingly began to spread rapidly, and continued to do so throughout the subsequent centuries, as can be seen in the graph and data shown on the next page, generated by *Gutenberg Auf!* from the Gutenberg corpus. Note that these are relative, rather than absolute, measures, and thus the particular volume of text available for a certain year will be less likely to skew the results. Having thus found evidence for the gradual increase in marked uses that we had predicted, we will continue by analysing uses of *if* and *even* during the time period preceding the emergence of the marked insufficiency *if*, namely during the 16<sup>th</sup> century, to test our predictions about the earlier stages of the emergence of the insufficiency *if*.

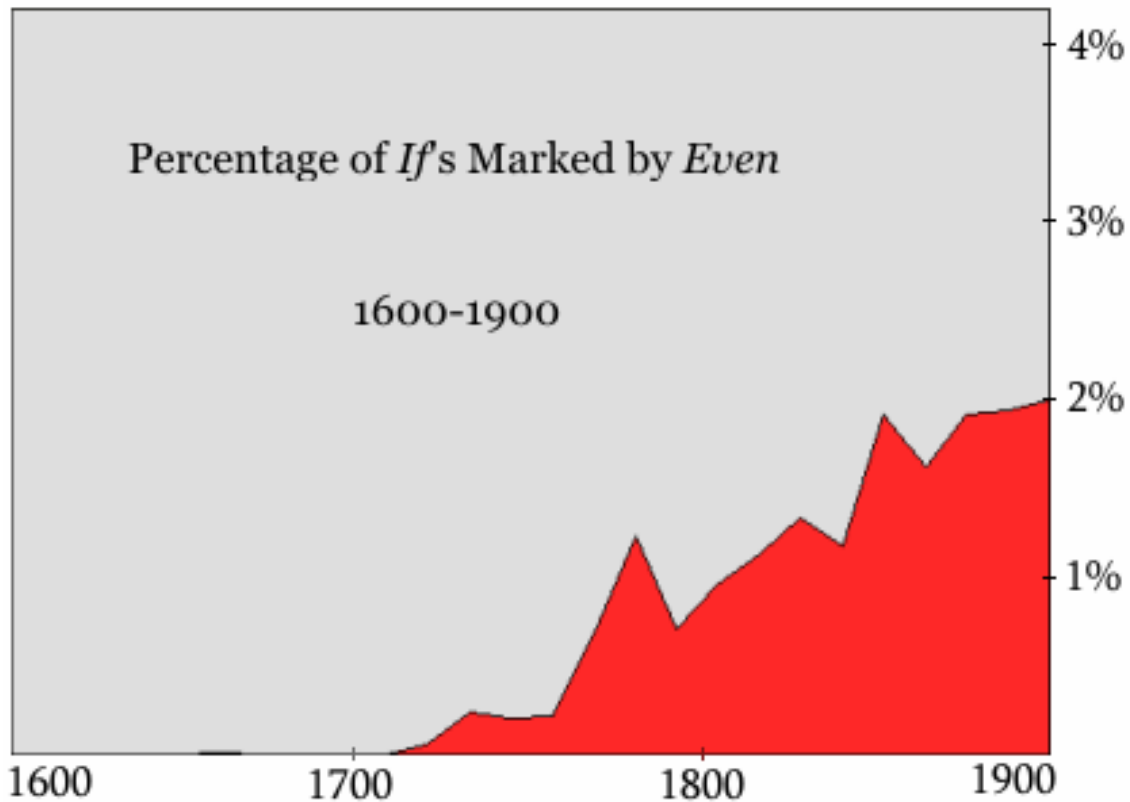
---

<sup>11</sup> Richard Hakluyt, *Voyager's Tales*, editor unknown; *vgrtl10.txt* in the Gutenberg corpus.

<sup>12</sup> Richard Hakluyt, *The principal nauigations, voyages, traffiques and discoueries of the English nation*, 1599; as available from EEBO at [http://gateway.proquest.com.proxy.lib.sfu.ca/openurl?ctx\\_ver=Z39.88-2003&res\\_id=xri:eebo&rft\\_val\\_fmt=&rft\\_id=xri:eebo:image:7116:844](http://gateway.proquest.com.proxy.lib.sfu.ca/openurl?ctx_ver=Z39.88-2003&res_id=xri:eebo&rft_val_fmt=&rft_id=xri:eebo:image:7116:844)

<sup>13</sup> This works out to 0.00316%

<sup>14</sup> This occurs in Charles Cotton's (1630-1687) translation of *The Essays of Montaigne*. However, even the existence of this *even if* is questionable: Cotton's translation was subsequently reworked by William Hazlitt in 1877, who remarked, in his preface to the new edition, that "[t]he besetting sin of both Montaigne's translators seems to have been a propensity for reducing his language and phraseology to the language and phraseology of the age and country to which they belonged". How marvellous that Hazlitt, then, was able to escape the idiom of his era!



VOLUMES AND RATIOS OF *EVEN IF* TO *IF*, 1380-1932

Dataset: Even If  
 Expression: \beven if\b

1380-1392: 0.0000000 (0 of 1469)	1764-1776: 0.0073823 (37 of 5012)
1440-1452: 0.0000000 (0 of 222)	1776-1788: 0.0054971 (23 of 4184)
1500-1512: 0.0000000 (0 of 369)	1788-1800: 0.0133303 (58 of 4351)
1512-1524: 0.0000000 (0 of 139)	1800-1812: 0.0072484 (105 of 14486)
1548-1560: 0.0000000 (0 of 76)	1812-1824: 0.0119661 (151 of 12619)
1572-1584: 0.0000000 (0 of 546)	1824-1836: 0.0098585 (147 of 14911)
1584-1596: 0.0000000 (0 of 2397)	1836-1848: 0.0121798 (494 of 40559)
1596-1608: 0.0000000 (0 of 10282)	1848-1860: 0.0161486 (1586 of 98213)
1608-1620: 0.0000000 (0 of 2499)	1860-1872: 0.0135663 (809 of 59633)
1644-1656: 0.0000000 (0 of 2451)	1872-1884: 0.0194707 (1843 of 94655)
1656-1668: 0.0000000 (0 of 133)	1884-1896: 0.0171723 (1248 of 72675)
1668-1680: 0.0000891 (1 of 11221)	1896-1908: 0.0191560 (1955 of 102057)
1680-1692: 0.0000000 (0 of 326)	1908-1920: 0.0207319 (2236 of 107853)
1692-1704: 0.0000000 (0 of 111)	1920-1932: 0.0220413 (896 of 40651)
1704-1716: 0.0000000 (0 of 2322)	
1716-1728: 0.0005102 (1 of 1960)	<i>For comparison:</i>
1728-1740: 0.0021433 (7 of 3266)	Google hits (Mar. 30, 2004):
1740-1752: 0.0015209 (8 of 5260)	<i>Even if</i> : 12.6 million vs. <i>If</i> : 487 million
1752-1764: 0.0031169 (6 of 1925)	(giving 2.59% marked)

## REFERENCES

Early English Books Online, <http://eebo.chadwyck.com/>

Jennings, R.E. and Schapansky, Nathalie. "Without: from separation to negation; A Case Study in Logicalization". *Proceedings of the 2000 Annual Conference of the Canadian Linguistics Association*. pp. 147-157.

Lycan, William. *Real conditionals*. Clarendon: Oxford, 2001.

Project Gutenberg, <http://www.gutenberg.net/>

Renaat, Declerck & Reed, Susan. *Conditionals: a comprehensive empirical analysis*. Gruyter: Berlin & New York, 2001.

Rissanen, Matti. "Computers Are Useful – For *Aught* I Know," in Colman, Fran (ed.), *Evidence for Old English*. J. Donald: Edinburgh, 1992. pp. 155-168.