J. Child Lang. **40** (2013), 438–468. © Cambridge University Press 2012
doi:10.1017/S0305000912000049

Structure of child and adult past counterfactuals, and implications for acquisition of the construction*

ALISON CRUTCHLEY
University of Huddersfield

(*Received* 13 *July* 2010 – *Revised* 6 *June* 2011 – *Accepted* 5 *January* 2012 – *First published online* 15 *March* 2012)

ABSTRACT

Children start producing *if p, q* conditionals relatively late. Past counterfactuals (PCFs), for example 'If she had shut the cage, the rabbit wouldn't have escaped', are particularly problematic for children; despite evidence of comprehension in the preschool years, children aged eleven are still making production errors in PCF structure (Crutchley, 2004). Working within a usage-based framework, the present study explores whether PCFs in the conversational component of the British National Corpus show structural similarities to the set of PCF structures produced by six- to eleven-year-old children in an elicitation task. Adult PCFs are found to be both rare in spontaneous conversation and very varied in structure. Low token frequency and high type frequency are hypothesized to account partly for children's late acquisition of the PCF construction. However, regularities in the use of subjects and verbs in adult PCFs are hypothesized to assist children's acquisition of the construction.

INTRODUCTION

Children have been found to acquire conditionals – of the form *if p, q* – relatively late. Reilly (1982, 1983, 1986) found in her studies of spontaneous productions that ''children do start to produce conditional sentences at about age 2½, but they do not fully control the entire conditional system until about 8 years of age'' (1986: 311). She noted that particular types of conditionals called 'present and past counterfactuals' were the latest emerging. Kuczaj and Daly (1979) found that children's ability to make explicit hypothetical reference began in the preschool years and was fairly well established by the age of seven, and commented that children were much more likely to produce future hypotheticals than to refer to past hypothetical events. Neither Reilly (1986) nor Kuczaj and Daly (1979)

[*] Address for correspondence: e-mail: a.crutchley@hud.ac.uk

https://doi.org/10.1017/S0305000912000049 Published online by Cambridge University Press

made a systematic study of the syntactic forms employed by children to convey conditional meanings, although Kuczaj and Daly noted that children used ''inappropriate forms'' when they first start making explicit hypothetical reference (1979: 569). Bowerman (1986) also observed that conditionals emerged relatively late, by comparison with other structures of similar complexity. She suggested that cognitive and pragmatic factors might be behind the late acquisition of conditionals by children, although she did not specify the precise nature of these factors. Therefore, most research into the acquisition of conditionals by children has focused on children's comprehension and use of conditional meanings, rather than conditional forms (see Crutchley, 2004, for an overview). There is still a dearth of information on how children acquire the syntactic structures used to express conditional meanings, and no clear theoretical explanation of why these structures – in particular, those which refer to past hypotheticals – appear so late in children's speech.

Reilly's and Bowerman's findings were supported by a study eliciting a particular subtype of conditional – the past counterfactual (PCF) – from a large group of six- to eleven-year-old children (Crutchley, 2004). This study found that control of the PCF was still emerging in this age range, speculating that the wide variety of forms produced by the children in response to the task could reflect variety in adult use of PCFs, and that this could contribute to their late acquisition.

The present article describes a study designed to explore this question. All the examples of PCFs that were in a large corpus of spontaneous conversational English were extracted and catalogued according to their syntactic form. Working within a usage-based framework, the present study relates the patterns of adult use of PCFs to those in Crutchley's (2004) study of children's elicited PCFs, and draws on notions such as type and token frequency to attempt an explanation of their late emergence in children's productive language.

The form of English conditionals

English conditionals encode a range of meanings and take a wide variety of syntactic forms. Even when analysis is restricted to *if p, q* conditionals, descriptive accounts such as Declerck and Reed (2001) detail a vast array of subtly different meanings encoded by a large number of similar structures.

Meaning differences are primarily expressed through the forms of the verbs used in the two clauses, and their relationship with one another. The effects of choosing different verb forms on the overall meaning of the sentence are well illustrated by the trio of conditional sentence types (which Declerck and Reed call ''canonical tense patterns'' (2001: 59)) typically

https://doi.org/10.1017/S0305000912000049 Published online by Cambridge University Press

encountered in teaching grammars and sometimes known as 'first, second and third conditionals':

(1) If you build it, they will come
(2) If you built it, they would come
(3) If you had built it, they would have come

Sentences (1) and (2) above are presented as referring to the future, with the verb forms encoding different levels of commitment on the part of the speaker to the reality or likelihood of the proposition in *p* being fulfilled. Sentence (3) is generally presented as encoding a special type of 'unlikelihood' – the proposition in *p* cannot be fulfilled, as the outcome of the event is already known. Sentences such as (3) have been called counterfactuals or 'impossible conditionals' (Schachter, 1971); the sentence in (3) is an example of a 'past counterfactual'. (Past counterfactuals contrast with present counterfactuals, which encode hypothetical/ impossible scenarios that are not linked to past events – for example, *If I ruled the world, every day would be the first day of spring*.)

Teaching grammars such as Thomson and Martinet (1986) concede that ''with each type certain variations are possible'' (p. 197), but in fact this model greatly simplifies the real-life usage of conditionals, where the *p* of one pattern may be 'mixed' with the *q* of another, as in:

(4) If I had failed, I would know it (example from Athanasiadou & Dirven, 1997)

Fillmore (1997a: 4), rather than distinguishing conditionals from one another on the basis of form, distinguishes three types of epistemic stance (ES) – positive, neutral and negative – and presents these as relating systematically to the choice of verb forms in conditional sentences. Counterfactual sentences encode negative epistemic stance, where ''the speaker assumes that 'P' is not true, where 'P' is a proposition derivable from (and preserving the polarity of) the form of the antecedent clause'' (p. 4). An advantage of Fillmore's approach is that it allows for a good deal of variation in the combination of verb tenses used in *p* and *q*. Examples of negative-ES sentences generated by Fillmore's model include:

(5) If you had eaten it, you would have died
(6) If you had joined the club, you would get invited to the reception
(7) If I were you, I would marry Louise

Fillmore's (1997a) account of conditional structure and meaning is based in a Construction Grammar approach. Beyond categorization by type of epistemic stance, Fillmore does not distinguish particular *if p, q* structures from one another. The past counterfactual is one of a number of possible

https://doi.org/10.1017/S0305000912000049 Published online by Cambridge University Press

negative-ES structures. A variation on this approach is to see the past counterfactual as a construction in its own right (albeit part of a set of related 'conditional' constructions). This approach assumes a clear relationship between a particular form (or set of forms) of the conditional and a particular meaning (or set of meanings). The rest of this article will take this approach.

Constructions

Construction Grammar approaches see constructions as the primary units of grammatical analysis. A loose definition of a construction is ''a pairing of form and meaning'' (Schönefeld, 2006: 18). Elsewhere, Fillmore provides the following definition:

> By grammatical construction we mean any syntactic pattern which is assigned one or more conventional functions in a language, together with whatever is linguistically conventionalized about its contribution to the meaning or the use of structures containing it. (Fillmore, 1988: 36, quoted in Schönefeld, 2006: 20).

Constructions occur in a range of forms, including idioms with more or less fixed formal and lexical characteristics (*it's raining cats and dogs*, *come a cropper*, *stark raving mad*); collocations such as '*blithering+idiot*' or '*consenting+adult*'; structures which are ''not complete runs, but have gaps in them'' (Fillmore, 1997b) such as '*It's* (*about*) *time you* [*brushed your teeth*]' or '*What's* [*the cat*] *doing* [*eating the baby's breakfast*]' (the '*What's* X *doing* Y' construction; Kay & Fillmore, 1999); and ''clausal patterns'' (Sag, to appear: 7), such as the auxiliary-initial pattern shown in examples like *Have I got news for you!* or syntactic 'frames' (called 'abstract syntactic constructions' by Tomasello, 2003) like Subject-Verb-Object 1-Object 2 (Goldberg, 2006). The past counterfactual construction falls into this last category.

The meaning and form of the PCF construction

Dancygier and Sweetser (2005) describe production and interpretation of conditionals in terms of Mental Spaces Theory:

> Our claim, then, is that … a conditional construction involves setting up a mental space (in the case of if-conditionals, this is the job of the if-clause), and requesting construal of something (in if-conditionals, the then-clause or main clause) within that space. (p. 18)

The past counterfactual construction can be seen as setting up a mental space in which an event in the past for which the outcome is already known

https://doi.org/10.1017/S0305000912000049 Published online by Cambridge University Press

is considered. The mental space derives all its features from the real world with the exception of this known outcome of the past event; instead an alternative outcome of the past event, or an alternative event, is construed within that space. A hypothetical effect of this alternative outcome or event is then posited.

The 'canonical' form of the past counterfactual is as in (3) above:

(3) If you had built it, they would have come

In the above example, *p* (the *if*-clause) has past perfect form (*had* + lexical verb). Fillmore (1997a) points out in reference to spoken American English that two non-standard variants can occur in *p*:

a. If you had have built it …
b. If you would have built it …

In addition, in some spoken varieties of British English, *have* and *of* are interchangeable in these two non-standard variants, so that

a. if you had of built it …
b. if you would of built it …

also occur. Thus the form of *p* consists of the past perfect, plus these four non-standard variants.

In some cases, the past event in *p* has an outcome that is not in the past. Fillmore (1997a) gives this example in (6) above:

(6) If you had joined the club, you would get invited to the reception

Therefore, the verb in *q* can take the form (*would* + *have* + lexical verb, or a non-standard variant) or (*would* + lexical verb).

Furthermore, continuous forms can also occur in the lexical verbs in both *p* and *q*, so the full set of PCF forms is as follows:

p	If she	had	shut the cage
		would have/of	been paying attention
		had have/of	
q	the fox	wouldn't have/of	eaten the rabbit
		wouldn't have/of	been eating the rabbitwhen we got home
	the rabbit	would still be alive	
		would still be running around	

To summarize the form of the construction, with parentheses indicating optional content:

p	***q***
if + had/would (+ have/of) + lexical verb	would (+ have) + lexical verb

https://doi.org/10.1017/S0305000912000049 Published online by Cambridge University Press

Children's acquisition of PCFs

The development of counterfactual reasoning. As noted above, it has been suggested that the late appearance of counterfactuals in children's spontaneous speech is partly to do with cognitive complexity. A number of psychological studies focus on comprehension of counterfactual meanings in young children. Beck, Robinson, Carroll and Apperley (2006) note in a review that ''typically-developing 4-year-olds can entertain counterfactuals with ease'' (p. 415), but that before this age counterfactual comprehension is problematic. Riggs, Peterson, Robinson and Mitchell (1998) showed that three- and four-year-old children coped better with tasks asking them to infer future hypothetical outcomes than tasks requiring them to reason about the state of the world now, had an earlier event not occurred (counterfactual reasoning), and this finding has been replicated by subsequent studies (e.g. Perner, Sprung & Steinkogler, 2004). (Note that this also chimes with Kuczaj & Daly's (1979) finding that children aged 2;6–5;6 produced future hypotheticals more frequently than past hypotheticals.) Robinson and Beck (2001) observe that ''to consider a counterfactual alongside the true state of affairs seems to stretch the mental resources of preschool children to their limits'' (p. 112), and some researchers suggest that processing load can account for children's difficulties with counterfactuals (e.g. Byrne, 1997). In this vein, German and Nichols (2003) found that five-year-olds had more difficulty with long causal chains (sequences of several events leading to an outcome) when interpreting counterfactuals, and suggest that long causal chains require more processing effort. However, Beck, Riggs and Gorniak (2010) point out that children aged three to four years may draw on real-world knowledge in the interpretation of many counterfactuals, giving the impression that they understand counterfactuality when they may not. They designed tasks which eliminated this intervening variable and found no clear relationship between length of causal chain and children's comprehension. Perner *et al*. (2004) compared 'simple' with 'complex' scenarios and found that three- to five-year-old children's performance on tasks requiring counterfactual reasoning was better for simple scenarios. They proposed that their 'complex' scenarios required children to make continual reference to the actual course of events in deriving outcomes, whereas simple scenarios allowed derivations to be made without this. Thus developments in memory for actual events and ability to cross-reference these events with hypothetical suggestions in deriving outcomes might play a part in the development of counterfactual reasoning.

General language level has been found to predict performance on counterfactual reasoning tasks. Riggs *et al*. (1998) found a positive relationship between receptive language level (as measured on the British

https://doi.org/10.1017/S0305000912000049 Published online by Cambridge University Press

Picture Vocabulary Scale (BPVS; Dunn, Dunn, Whetton & Burley, 1997)) and performance on tests of counterfactual reasoning in three- and four-year-old children. Guajardo and Turley-Ames (2004) found a positive relationship between three- to five-year-old children's level on a broader measure of receptive language (the Test of Auditory Comprehension of Language – Revised (TACL-R; Carrow-Woolfolk, 1985)) and performance on counterfactual reasoning tasks. In one of the few studies to consider the linguistic form of counterfactuals in relation to comprehensibility, Perner *et al*. (2004) found that counterfactuals with a subjunctive form in *p* were no more difficult for three- to five-year-old children to understand than those without, and counterfactuals with two subjunctives (one in *p* and one in *q*) were no more difficult than those with a subjunctive only in *p* (although this research was conducted in German rather than English).

A number of factors, then, may potentially play a part in making counterfactual reasoning difficult for young children: processing load, memory, real-world knowledge and linguistic ability. However, the relationship between comprehension and production of counterfactuals is not clear. Assuming that Perner *et al*.'s (2004) results are applicable to English as well as to German, the linguistic structure of counterfactuals may not make a difference to children's comprehension. Nevertheless, Kuczaj and Daly (1979) found children using 'inappropriate forms' to express hypotheticals. While studies of child language acquisition have found repeatedly that production lags behind comprehension, the studies summarized here indicate that children may be able to comprehend counterfactuals as young as three or four years of age, but still be producing them inconsistently at the age of eleven. Therefore, it seems possible that there is a partial dissociation between linguistic and cognitive abilities in relation to the production of counterfactuals. How do children learn to express counterfactuality in an adult-like way? Crutchley (2004) made some initial hypotheses about the productive acquisition of PCFs based on an elicitation study. These form the basis for the aims of the present study.

Crutchley (2004) used picture stimuli to elicit PCFs from a stratified sample of 799 children aged six to eleven.[1] Some 41 percent of children's responses had the target (i.e. 'canonical' PCF) structure, of the type '*if* + *had* + lexical verb, *would* + *have* + lexical verb', as in '*If she had shut the cage, the rabbit wouldn't have escaped*'. Crutchley found that children's ability to produce the target structure improved with age, but that even in the oldest age group, a proportion of children still produced structures other than the

[1] The data described in Crutchley (2004) and in Study 2 in the present article were collected during the standardization phase of the development of the Assessment of Comprehension and Expression 6–11 (ACE 6–11) (Adams, Cooke, Crutchley, Hesketh & Reeves, 2001).

https://doi.org/10.1017/S0305000912000049 Published online by Cambridge University Press

target in response to the task. She concluded that control of PCF structure was still developing in this age group (p. 230), and noted the absence of clear age-related 'stages' in production of PCFs in response to this task (p. 230). Developmental progression was clear in the dataset, but this consisted mainly of increased usage of adult-like PCF forms with increasing age, and decreased usage of non-conditional structures such as picture descriptions and structures using modals (e.g. *She should have shut the cage*). Children did produce other conditionals (e.g. *If she shut the cage, the rabbit wouldn't escape*) but there was no evidence that younger children were systematically substituting other conditionals for PCFs as a developmental stage.

This observation that children do not progress through clear stages of using other conditional forms before productively controlling the PCF suggests that there is no clear relationship between children's acquisition of non-PCF conditionals and their acquisition of the PCF construction. Instead, it suggests that children gradually approximate adult-like usage of the whole PCF construction at once, rather than explicitly relating it to other conditional forms.

Perhaps the most surprising finding from this study was the sheer variety of structures produced by the children in response to the task. Although five structures accounted for 74% of the responses, children produced thirty-six different 'conditional type' structures in total (Crutchley, 2004: 219). Despite this variety, children's responses demonstrated their tacit understanding of various aspects of the PCF construction. Some 84% of responses were of the form *if p, q*; the verb tenses in the two clauses were generally different; the tense in *p* nearly always 'preceded' that in *q*, showing an understanding of the 'compatibility relations' (Fillmore, 1997a: 3) of the verb form in *p* with that in *q*; and the majority (77%) of clause combinations were logically coherent (e.g. in explaining a cause-and-effect relationship between the two clauses). Logical incoherencies were particularly rare in older age groups. Additionally, reanalysis of the data from Crutchley indicates that 77% of children's responses included a past perfect verb form in *p* or *q*, characteristic of the 'canonical' PCF construction as outlined above. Furthermore, Crutchley argued that many of the 'non-adult-like' structures that children produced in that study still reflected tacit understanding of ways in which verb tenses could be used to convey different types of meaning (e.g. habitual actions), even if that understanding was not realized in an adult-like production.

Crutchley (2004) hypothesized that the great variability in syntactic form observed in children's productions might reflect variability in adult usage. Usage-based accounts of language acquisition propose that frequency of occurrence of a construction in the ambient language has an important relationship with the ease and speed of children's acquisition of that

https://doi.org/10.1017/S0305000912000049 Published online by Cambridge University Press

construction (Clark & Kelly, 2006). Clark and Kelly explain the mechanism thus:

> Frequency of exposure should help children set up conceptual representations for forms, linked to whatever meanings they have identified so far. Greater frequency of exposure should make for greater cognitive entrenchment, with the result that more frequent form-meaning pairs are more readily recognised and more easily retrieved for production. This points still more strongly to the need for detailed studies of usage to establish just which constructions children are exposed to, their communicative functions, and their relative frequency. To date, we know relatively little about any of these (but see Cameron-Faulkner *et al.*, 2003), and we know virtually nothing about the amount of variation across adults in which constructions they might favour for a particular function and so use more frequently. (p. 6)

Clark and Kelly summarize a number of studies showing that children are sensitive to frequencies in the input from an early age (e.g. Goldberg, Casenhiser & Sethuraman, 2005; Huttenlocher, Smiley & Charney, 1983; Tomasello, 1992). Frequency of exposure to a construction can take two different forms, whose effects, as Bybee and Thompson (2000) point out, are partly separable: token frequency and type frequency. Bybee and Thompson define the two terms as follows: ''Token frequency is the count of the occurrence in texts of particular words ... or of specific phrases ... Type frequency, on the other hand, counts how many different lexical items a certain pattern or construction is applicable to'' (p. 378).

Corpus-based studies of spoken language, while relatively rare compared to studies based on written language, suggest that conditionals of all types are rare in spontaneous speech (Athanasiadou & Dirven, 1997; Declerck & Reed, 2001). This suggests that children will hear relatively few instantiations of PCFs (low token frequency). Second, the type frequency is very high: PCFs involve a large number of different lexical items and a number of structural variations. High type frequency plus low token frequency ensures that children are unlikely to hear repeated instantiations of the construction involving the same lexical items. This means that in order to recognise and internalize the construction as such, children need to be able to abstract away from these individual instantiations what they have in common – the syntactic form. As noted above, Crutchley (2004) found some evidence that children are, in fact, able to make some of these abstractions, even before they control the PCF construction in an adult-like way.

Nevertheless, even where there is great variability in structure, subtle frequency effects may exist. These can be uncovered with analysis of large

https://doi.org/10.1017/S0305000912000049 Published online by Cambridge University Press

datasets. Goldberg (2006) summarizes the results of several studies looking at the verbs used by caregivers in particular constructions. She notes that, although some constructions occurred with a wide range of verbs, caregivers often used one verb much more frequently than the others in a particular construction. For example, Goldberg *et al*. (2005) looked at the Subject-Verb-Oblative construction (e.g. *The milk goes in the fridge*) in a database of mother–child interactions (children at age 1;8 and 2;4 (Bates, Bretherton & Snyder, 1988)). They found that the construction was used with a total of thirty-nine different verbs. Despite this variety, the verb 'go' was used in 39 percent of instantiations of this construction (136 out of 353 instantiations in the data). 'Go' is, of course, a high-frequency verb, but statistical adjustment for this general frequency effect did not account for this large proportion. Similar effects were found for other verbs in different constructions in Goldberg *et al*.'s study.

Goldberg suggests that this 'skewed' input helps children to acquire some abstract syntactic constructions as they hear them disproportionately frequently with the same verbs. She also suggests that the verbs that are used frequently in these cases are those which (a) are high frequency in the input, so are likely to be accessible to the child and (b) exemplify the typical meaning of the construction. For example, in Goldberg *et al*.'s (2005) study, the Subject-Verb-Object 1-Object 2 construction occurred most frequently with the verb 'give', exemplifying the typical meaning of 'someone causing someone to receive something' (Goldberg, 2006: 77). These instances can then form the basis of generalizations about the meaning and form of the construction. If effects such as these are also found in adult spontaneous usage of PCFs, this could be hypothesized to be assisting children in the acquisition of the PCF construction.

Frequency effects may apply to other components of constructions as well as verbs. Cameron-Faulkner, Lieven and Tomasello (2003) looked at the beginnings of utterances addressed to children. They found a great number of repeated 'item-based phrases' such as *Are you …* , *I'll …* , *Look at …* , etc., and noted that many of these were replicated by the children in their early utterances. They also found a preponderance of pronouns used as subjects. If the PCF is thought of as an abstract syntactic frame, an infinite number of lexical items could occur in the 'slots' for 'lexical verb' and 'subject' in the two clauses, and if this is reflected in actual usage, it presents a significant challenge for children in noticing and abstracting away the grammatical categories that make up the construction. However, if adult usage is actually less varied than this abstract model suggests, and particular lexical items occur frequently within PCFs, the token frequency of these part-constructions could feasibly assist children in acquiring the basis of the entire construction. Tomasello characterizes

https://doi.org/10.1017/S0305000912000049 Published online by Cambridge University Press

this as a process of "*functionally based distributional analysis*, in which the learner groups together into categories those linguistic items that function similarly – that is, consistently play similar communicative roles – in different utterances and constructions" (2003: 145; emphasis in original).

The present study aims to investigate the nature of spontaneous adult PCFs with these questions in mind.

THE PRESENT STUDY

Two studies were carried out to investigate adult use of PCF forms in spontaneous speech, and possible relationships with children's acquisition of the construction. Study 1 analyzed adult usage of PCFs in a large corpus of spontaneous conversational data; Study 2 recoded and reanalyzed existing data from a study eliciting PCFs from a large group of six- to eleven-year-old children (Crutchley, 2004).

Study 1 aims

The aims of Study 1 were to use a large corpus of spontaneous adult conversational data to answer the following questions:

1. How frequent are PCFs in adult conversational data?
2. How much do PCF structures vary in adult conversational data?
3. Which PCF structures are most frequent in adult conversational data?
4. Which verbs are most frequent in PCFs in adult conversational data?
5. Which subjects are most frequent in PCFs in adult conversational data?
6. How closely do the PCF structures used in adult conversational data reflect the ideal characterization of PCF form outlined earlier?

Study 2 aims

The results from the corpus analysis then fed into discussion of the following questions relating to children's acquisition of PCFs:

1. How closely do the PCF structures produced in adult conversational data match those produced by children in Crutchley's (2004) study?
2. How closely do the verbs produced in PCF structures in adult conversational data match those produced by children in Crutchley's (2004) study?
3. How closely do the subjects produced in PCF structures in adult conversational data match those produced by children in Crutchley's (2004) study?

https://doi.org/10.1017/S0305000912000049 Published online by Cambridge University Press

STUDY 1

METHOD

In order to investigate adult spontaneous usage of PCFs, a number of searches were made of the conversational component of the British National Corpus (BNC) (British National Corpus, 2007). The BNC was chosen as it is a large, representative corpus of British English containing a substantial proportion of spontaneous spoken language. A large corpus such as the BNC allows insights to be drawn from overall proportions of different structures in the data, including structures that occur rarely in spontaneous speech. Effects such as the disproportionate use of certain verbs in particular PCF structures may only be visible with a very large dataset. The conversational data for the BNC were collected by a demographically representative sample of 124 adults (age 15+) drawn from across the UK and representing all age groups and social classes, who wore recording devices and recorded their naturally occurring conversations with a large number of conversational partners over a two- to seven-day period (Burnard, 2000: 12). The total number of speakers in conversations is not given, but Burnard states that ''the total number of participants in all conversations was well in excess of a thousand'' (p. 14). The total number of words in the BNC conversational component is given as 3,919,712 (p. 14).

Searches were conducted using the WordSmith Concord tool (Scott, 2004). This tool allows the user to specify a word to be searched for and one context word (within up to 25 words to the left or right of the search word). In order to reflect the range of structures potentially being used to express past counterfactual meanings, the initial searches were designed to return any bi-clausal *if p, q* sentence which had EITHER the following form in *p*:

if + had + lexical verb

OR the following form in *q*:

would + have + lexical verb.

This then meant that examples such as

(8) He'd be on his back if he'd had that amount of drink [kdn037][2]
(9) if there is an easy way of doing it the Conservatives would have found out and done it [kbc093]

were included in the results.

[2] Data cited herein have been extracted from the British National Corpus, distributed by Oxford University Computing Services on behalf of the BNC Consortium. All rights in the texts cited are reserved. As requested by the British National Corpus (2007), citations from the BNC are labelled with the text identifier (3-character code) and sentence number (3-number code). So [kdn037] refers to text kdn, sentence number 37.

https://doi.org/10.1017/S0305000912000049 Published online by Cambridge University Press

In addition, searches needed to take into account that *have* and *of* are pronounced very similarly in connected speech and may be transcribed interchangeably, and that *had* and *would* can both be truncated to *'d*.

The following searches were performed:

if + had
if + 'd
would + have
would + of
would + 've

These searches returned a large number of concordances, which were then examined individually and rejected if they did not fit the structural criteria outlined above. For example, the search for '*if* + *had*' returns

(10) I just wondered if you had anything planned for the weekend [kbv032]

which is not a PCF and was rejected. The structures of *p* and *q* in each PCF were then coded numerically: structures were coded with numbers in the order in which they appeared in the data. So '*if* + *had* + lexical verb' was coded 1, '*if* + *had* + *have* + lexical verb' was coded 2, and so on.

In order to explore possible frequency effects from recurrent verbs and subjects in PCFs, I noted which verbs were used in the two clauses of each PCF, and the structures in which different verbs were most commonly used, in case particular verbs tended to be used in particular structures. I also noted the subjects which occurred in the two clauses of each PCF, in case particular subjects were frequent enough to suggest that item-based constructions might be formed around them.

Reliability

As recommended by Lombard, Snyder-Duch and Bracken (2002), a subset of the PCFs from the above corpus search were coded independently for structure by a researcher with experience of grammatical analysis, who had been trained on a small sample of structures taken from the BNC. For the inter-rater reliability sample, fifty PCFs were selected using a web-based random number generator (www.random.org). Fifty PCFs represent 12·9 percent of the total number of PCFs found in the BNC (see 'Results', below). Krippendorff's alpha was selected as the most appropriate means of measuring inter-rater agreement where multiple coding categories are used (Lombard *et al.*, 2002; Krippendorff, 2007). Two separate calculations were made, for *p* and *q*. For *p*, $_{\text{nominal}}\alpha = 0{\cdot}893$; for *q*, $_{\text{nominal}}\alpha = 0{\cdot}936$. These results indicate very good agreement between the two raters (Krippendorff, 2004: 241–43).

https://doi.org/10.1017/S0305000912000049 Published online by Cambridge University Press

RESULTS

Structure of adult PCFs

The BNC searches outlined above, coupled with manual examination of each item, returned a total of 389 PCF examples. Burnard (2000: 13) indicates that the BNC conversational corpus contains a total of 610,563 utterances. Thus, 389 PCFs represent 0·06 percent of utterances in conversations in the corpus. Within these 389 examples, 43 different syntactic structures were found. (A list of these with examples from the data can be found in the 'Appendix'.) Many of these structures occurred only once or twice in the data.

The PCF form was characterized above as:

p	**q**
if + had/would (+ have/of) + lexical verb	would (+ have) + lexical verb

Two hundred and thirty-six (60·7%) of the PCFs in the BNC have the above structure in *p* and '*would* + *have* + lexical verb' in *q*. Over half of these have the 'canonical' PCF structure, '*if* + *had* + lexical verb … *would* + *have* + lexical verb' (this was the most commonly occurring structure in the dataset, accounting for 34·7% of the PCFs found in the corpus). This set also contains the second most commonly occurring structure in the corpus, '*if* + *had* + *have/of* + lexical verb … *would* + *have/of* + lexical verb', which accounted for nearly a quarter of the PCFs in the corpus (23·4%). This set contains a further four structural variants, including:

(11) if we would have gone somewhere else and had coffee and a cake it would of cost you one fifty each [kc8036] ('*if* + *would have/of* + lexical verb … *would* + *have/of* + lexical verb')

(12) if our Margaret had er been working, I honestly think she would have left Pete. [kb1036] ('*if* + *had* + past cont … *would* + *have/of* + lexical verb')

In addition, thirty-seven (9·5%) of the BNC PCFs have the above structure in *p* and '*would* + lexical verb' in *q*, e.g. (8) above:

(8) He'd be on his back if he'd had that amount of drink [kdn037]

This set contains a further eight structural variants.

In total, then, 273 (70·2%) of PCFs in the BNC conform to the above characterization of PCF form. However, the third most commonly occurring structure in the corpus does not conform to this characterization. It has '*if* + past simple' in *p* and '*would* + *have/of* + lexical verb' in *q*, as in:

(13) if I came back half an hour later, she'd still of been there [kdl028]

Fifty-three of the PCFs in the corpus (13·6%) have this structure. This structure was also produced by the children in Crutchley (2004); indeed, it

https://doi.org/10.1017/S0305000912000049 Published online by Cambridge University Press

was the second most frequently occurring structure in that study (after the 'canonical' PCF structure), accounting for 16% of the data. Children produced responses such as *If he wasn't late for the bus, he wouldn't have missed it*, which seemed to encode purely counterfactual meanings. Crutchley (2004) asked a small group of adults to rate the grammaticality of this (and other) PCF-like structures. The raters could not agree on whether this structure was grammatical or not. Despite this, this structure is frequent among PCFs in the BNC (58 examples; 14·9% of the total number of examples).

Just under half of the BNC examples with the structure '*if* + past simple … *would* + *have* + lexical verb' use the structure to encode a past counterfactual meaning (21/53). For example:

(14) If my children came home and said to me that they got the cane at school I'd have just said well you must have deserved it [kb7003]
(15) if they took my wages into consideration they would have let us buy next door even [kb7098]
(16) if I lived with him first, I would never of married him [kc9050]

In all the above examples, the past simple in *p* could be replaced with the past perfect without a change in meaning. It therefore seems possible that, at least for some speakers or perhaps in some contexts, this structure is part of the PCF construction.

However, in thirty out of fifty-three examples, this structure is used to encode a general state of affairs in *p*, while referring back to a specific event in *q*, for example as evidence to support the proposition in *p*:

(17) if she were your daughter Jackie what would you have thought? [kbb027]
(18) I think if he wanted you he would have come to you, don't you? [kbc087]
(19) if he thought there was something up he would of sent you for an x-ray, or a blood test [kbr020]
(20) we wouldn't have invited him if we weren't sympathetic would we? [kb0096]

Therefore, although this structure is not unequivocally part of the PCF construction, it does seem that it is sometimes used to convey a PCF meaning.

Fifty-eight PCF-like forms (14·9% of the occurrences in the BNC) do not fit into the model outlined above, and do not have the structure '*if* + past simple … *would* + *have* + lexical verb'. Of these, twelve reflect possible transcription errors (e.g. *if I'd known I'd gone down there* [kct017] might have been a mistranscription or mishearing of [aɪdəgɒn] as [aɪdgɒn]). Interpretation of the other occurrences will depend on theoretical

https://doi.org/10.1017/S0305000912000049 Published online by Cambridge University Press

TABLE 1. *The most frequent verbs in adult PCFs*

	p		q	
Verb	Frequency	Percent	Frequency	Percent
be	70	18	100	25·7
have	36	9·3	27	6·9
know	27	6·9	4	1
go	21	5·4	19	4·9
get	14	3·6	22	5·7

orientation. Accounts of conditionals based on corpus data tend to dismiss these as 'errors' (e.g. Declerck & Reed, 2001; Athanasiadou & Dirven, 1997), akin to Chomskyan performance errors, where speakers reformulate midway through an utterance or forget the form they were planning to use. However, on closer inspection, it seems that some of these usages may be motivated by the speaker's desire to create a non-prototypical meaning. For example, some speakers seemed to use PCF-like forms to heighten the contrast between the speaker's evaluation of the propositions in *p* and *q*. In some of these examples, the past perfect/past conditional did not necessarily convey past time:

(21) if you hadn't got it by the Tuesday, to let us know on the Wednesday aye, next week mm [kss051]

Here, use of the past perfect in *p* ('hypothetical backshift'; Dancygier, 1998: 39) increases the 'distance' between the verb forms in the two clauses, emphasizing the speaker's negative epistemic stance (Fillmore, 1997a: 4) towards the proposition in *p* and his/her relatively positive epistemic stance towards the proposition in *q* (compare *If you haven't got it by the Tuesday* …). Examples such as these indicate that PCF-like forms are used flexibly by a small proportion of speakers to convey non-PCF meanings.

Verbs in adult PCFs

A great variety of verbs were used in the 389 PCFs in the data. In *p*, 119 different verbs were used; in *q*, 117. However, certain verbs were more common than others. A small number of verbs occurred in more than 5 percent of utterances (see Table 1).

As can be seen in Table 1, the verb 'be' occurred much more frequently than any other verb: 70 times in *p* (18%) and 100 times in *q* (25·7%). This suggests that structures like '*If* X *had been* Y' or 'X *would have been* Y' occur frequently. Note that, while 'go' and 'get' occurred frequently in both clauses, 'know' occurred 27 times in *p*, but only 4 times in *q*. '*If* X *had known* Y' therefore occurred frequently, but not 'X *would have known* Y'.

https://doi.org/10.1017/S0305000912000049 Published online by Cambridge University Press

Despite the similarity in overall verb frequencies across the two clauses, combinations of verbs in the two clauses were less predictable. The most commonly occurring pairing was '*if*+be … *would*+be', which occurred 18 times (4·6% of PCFs), as in the following examples:

(22) Er, see if it had been a bit more local it wouldn't have been so bad [kd8003]
(23) Would it have been better if it had not been a corner ball? [kbk076]

No other pairs of verbs appeared frequently in PCFs, suggesting that *p* and *q* tended to pattern separately.

Goldberg (2006) suggested that if particular verbs occurred with great frequency in particular structures, this might assist children in noticing a 'typical' meaning for the structure. Therefore, the data were examined to see if particular verbs occurred frequently in the most common PCF structures. For example, the 'canonical' structure, '*if*+*had*+lexical verb … *would*+*have*+lexical verb', accounted for 35% of PCFs in the data. Thus, if verbs are evenly distributed across different PCF structures, we would expect common verbs to occur in roughly this proportion in this structure (i.e. 35% of instances of 'be', 'have', 'do', etc. should occur in examples of this particular structure). The next most frequent structure, '*if*+*had have*+lexical verb … *would*+*have*+lexical verb', accounted for 23% of PCFs in the data, so we would expect roughly 23% of the instances of each of the common verbs to be found in this structure, and so on for the other frequent structures.

Overall, this is indeed what was found. However, there were some anomalies. For example, 50% of instances of 'do' and 48% of instances of 'know' in *p* occurred in '*if*+*had*+lexical verb … *would*+*have*+lexical verb', where around 35% would be expected for both. Some 37% of instances of 'know' in *p* occurred in the structure '*if*+*had have*+lexical verb … *would*+*have*+lexical verb', where around 23% would be expected. See Table 2 for a summary.

The unexpected frequency of certain verbs in certain structures suggests that repeated combinations such as '*If* A *had done* B, C *would have* D', or '*If* A *had've known* (B), C *would have* D' may occur in the data.

Subjects in adult PCFs

Some 89·7% of subjects in *p* and 88·9% of subjects in *q* were pronouns. This is in line with existing corpus research, which indicates that full-NP subjects are very rare in conversation (see Francis, Gregory & Michaelis, 1999, for a review). Certain pronouns were particularly frequent. In both *p* and *q*, the most commonly occurring subject by far was *I*, appearing 107 times in *p* (27·5%) and 127 times in *q* (32·6%), suggesting that PCFs are often used to

https://doi.org/10.1017/S0305000912000049 Published online by Cambridge University Press

TABLE 2. *Verbs which commonly occur in particular adult PCF structures*

Structure (percentage of total PCFs which have this structure)	Common verbs in each structure					
	p			q		
	verb	freq	%	verb	freq	%
if + had + lexical verb, would + have/of + lexical verb (35%)	do	9/18	50			
	know	13/27	48			
if + had have/of + lexical verb, would + have/of + lexical verb (23%)	know	10/27	37	do	5/14	36
				get	7/22	32
				have	10/27	37
if + past simple, would + have/of + lexical verb (14%)	be	22/70	31	do	3/14	21
				go	4/19	21
				say	3/12	25

talk about one's own situation or events affecting oneself. The next most frequent subject was *you* (71 examples in *p*, 18·3%; 50 examples in *q*, 12·9%). Personal pronouns (*I*, *we*, *you*, *he*, *she*, *they*) accounted for 79·7% of subjects in *p* and 71·5% of subjects in *q*.

Some combinations of subjects in *p* and *q* were particularly frequent. Predictably (as *I* was such a frequently occurring subject), PCFs that had *I* as the subject of both *p* and *q* were by far the most common combination (39 occurrences, 10% of all PCFs). The next most frequent combinations were '*you* + *you*' (28 occurrences, 7·2%). This could indicate that combinations such as '*If I'd* X … , *I'd have* Y …' and '*If you'd* X, *you would have* Y' occur repeatedly in spontaneous conversation.

In order to investigate whether frequent subjects and frequent verbs might occur together and produce more complex repeated combinations, I noted the verbs that occurred with frequent subjects in *p* and in *q*. The most frequent combination, '*it* + 'be'', occurred 39 times in *q* (10% of *q* clauses).

(24) It would have been nice if he'd waited for there to be a gap in the traffic [ke3009]

(25) It would have been fantastic if they'd had the cash to just get on wouldn't it [ke6060]

The majority of these *q* clauses (25/39, 64·1%) had the structure '*would* + 'be' + comparative adjective/NP' (e.g. *better*, *all right*, *nice*, *nicer*, *different*, *good*, *fine*, *fantastic*, *the same*).

SUMMARY OF FINDINGS FROM STUDY 1

PCFs are relatively rare in spontaneous conversation (accounting for around 0·05% of utterances in the BNC corpus), and variable in structure:

https://doi.org/10.1017/S0305000912000049 Published online by Cambridge University Press

forty-three different PCF structures were found in the corpus. Some PCF structures were much more common than others: four structures accounted for 77 percent of the PCFs in the corpus. Nevertheless, even the most commonly occurring structure – the 'canonical' PCF – only accounted for around one-third of examples in the corpus. Thus, variety in structure is the norm in conveying past counterfactual meanings in spontaneous conversation. Past counterfactuality may also be encoded in a form that is not associated with this meaning in existing descriptive accounts of conditionals: '*if* + past simple … *would* + *have* + lexical verb' (although this form does not pair completely reliably with this meaning). PCF-like forms are also occasionally used to convey non-PCF meanings.

A great variety of different verbs were used in PCFs in the BNC. Nevertheless, some verbs were used very frequently, e.g. 'be'. Patterns of verb use in *p* and in *q* were not always the same – for example, 'know' occurred almost exclusively in *p* – suggesting that general frequency of verbs in spontaneous language could not entirely explain verb behaviour. Certain verbs were also found to occur more frequently in particular PCF structures, suggesting that combinations such as '*If* A *had've known* B, C *would have* D' may occur reasonably frequently.

As expected in conversational data, most subjects were pronominal in form in both *p* and *q*. In both *p* and *q*, the most frequent subject was *I*, suggesting that PCFs are often used to talk about the speaker's own situation. Certain verbs and subjects also occurred frequently in particular PCF structures, suggesting that combinations such as '*If* … , *it would have been* [evaluative expression]' may occur relatively frequently.

STUDY 2

Method

Crutchley (2004) reported the responses of a large stratified sample of six- to eleven-year-old children to picture stimuli designed to elicit PCFs. Before responding to the test stimuli, children did one practice item showing a girl with an empty homework book, imagining herself doing her homework. Her teacher is waving a finger at her angrily. The fieldworker explained the task to the child, and modelled the 'canonical' PCF: *If she had done her homework, she wouldn't have been told off by the teacher*. The 799 children produced two responses each. The first stimulus picture featured a rabbit running out of an open cage with a girl attempting unsuccessfully to catch it, and imagining herself locking the cage with the rabbit inside. The second stimulus picture depicted a boy in school uniform watching a bus pulling away, and imagining himself running next to it and making for its open door.

https://doi.org/10.1017/S0305000912000049 Published online by Cambridge University Press

Study 2 re-examined the data from Crutchley (2004), noting frequencies of different verbs and subjects in the children's responses, as these were not examined in the original analysis. Frequencies of different conditional structures, verbs and subjects in the children's data could then be compared with findings from the BNC study of adult conversation reported above.

Exclusions from the analysis

Forty-four of the 'bus' responses and fifty-nine of the 'rabbit' responses had unknown verbs as the fieldworker had not noted the actual response but simply noted that it had the target structure. These responses were excluded from subsequent analyses. Responses which did not have a 'conditional' form were also excluded from analysis (e.g. picture descriptions, modal structures such as *She should have shut the cage*). The total number of remaining responses analysed was 632 for the rabbit stimulus, and 608 for the bus stimulus.

Phrasal verbs (verb + particle combinations with non-compositional meanings) were initially coded individually. However, some base verbs occurred in combination with several different particles (e.g. *get in*/*get into*/*get on the bus*) and expressed closely related meanings. Therefore verbs of this type were grouped into single categories according to the base verb. All subsequent calculations use these categories.

RESULTS

Children produced thirty-six different conditional structures. Four structures each occurred more than 5 percent of the time. These will be discussed below in relation to the findings from Study 1 (Table 4).

Verbs in children's PCFs

A total of sixty-five different verbs occurred in children's responses to the two stimuli. Forty-eight different verbs were used in *p*; forty-one different verbs were used in *q*. Twenty verbs were used in both *p* and *q*; seven verbs were used in all four clauses. Six verbs each accounted for more than 5 percent of responses: these are listed in Table 3.

Some frequent verbs reflected aspects of the stimuli. 'Miss' occurred almost exclusively in *q* of 'bus' responses, as children described the boy missing the bus. 'Lock' occurred only in *p* of 'rabbit' responses, as children produced responses like *If she had locked the cage*. … However, some verbs were used frequently across both stimuli. The most frequent verb overall, 'run', appeared almost equally frequently in 'bus' and 'rabbit' responses. Children described the rabbit running away, and the boy running for the

https://doi.org/10.1017/S0305000912000049 Published online by Cambridge University Press

TABLE 3. *Most frequent verbs in the children's data*

	Responses to 'rabbit' stimulus – frequency		Responses to 'bus' stimulus – frequency		Total across both stimuli	Percent across both stimuli
	p $n=608$	q $n=608$	p $n=632$	q $n=632$	p and q $n=2480$	p and q
run	1	279	272	5	557	22·5
get	5	144	118	104	371	15·0
miss			9	203	212	8·6
be + ADJ/ NP	1	11	100	54	166	6·7
catch	2	4	19	133	158	6·4
lock	142	0	0	0	142	5·7

TABLE 4. *Comparison of most frequent PCF structures in adult and child data*

	Adult data			Child data		
PCF structure	frequency ($n=389$)	%	rank	frequency ($n=1598$)	%	rank
if + had + lexical verb, would + have/of + lexical verb	135	34·7	1	651	40·7	1
if + had + have/of + lexical verb, would + have/of + lexical verb	91	23·4	2	133	8·3	3
if + past simple, would + have/of + lexical verb	53	13·6	3	248	15·5	2
if + had + lexical verb, would + lexical verb	23	5·9	4	43	2·7	6
if + would have/of + lexical verb, would + have/of + lexical verb	6	1·5	5	95	5·9	4

bus. Two other frequent verbs were used across clauses and response types. ''Get' + particle' combinations occurred in 15 percent of clauses, and ''be' + comparative ADJ/ NP' occurred in 6·7 percent of clauses.

Subjects in children's PCFs

Overall, 87% of subjects in *p* and 62% in *q* were pronouns. There was some variation in responses to the two stimuli: 95% of subjects in *p* in 'rabbit' responses were *she*; 73% in *p* in 'bus' responses were *he* (children also referred to the subject in this clause as *the boy* (12%) and *the man* (6%)). The subject in *q* of 'rabbit' responses is the only one to be predominantly realized as a full NP: *the rabbit* makes up 63% of children's responses in this clause. The majority of children's 'rabbit' responses (353/632, 56%) had *she* in *p* and *the rabbit* in *q*, as in *If she had shut the cage, the rabbit wouldn't have escaped*; the next most common combination of subjects was '*she* + *she*'

https://doi.org/10.1017/S0305000912000049 Published online by Cambridge University Press

(72/632 responses, 11%). In contrast to this, the majority of the children's 'bus' responses had the same subject (*he*) in *p* and *q*, in formulations such as *If he hadn't missed the bus, he wouldn't have been late for school* (408/608 responses, 67%). Whether the reference is switched between *p* and *q* is therefore reflected in choice of subject NP: where the subject is the same in both clauses, no nominal subjects are used. Some 28 children used *I* as a subject in *p*, and 18 in *q*. While these numbers are low, it is interesting that children used first person pronouns at all, given that the task was a picture description and the practice item modelled a response with *she*.

SUMMARY OF RESULTS FROM STUDIES 1 AND 2

While both adults and children showed a great deal of structural variety in the PCFs they produced, there were similarities between the two studies in the structures used most frequently. Four structures accounted for more than 70 percent of the responses in both Study 1 and Study 2. As can be seen in Table 4, the most frequently occurring structures corresponded closely in the two datasets. The most frequently occurring structure in both studies was the 'canonical' PCF, '*if*+*had*+lexical verb … *would*+*have*+lexical verb'. The next two most common structures in the BNC appear in reverse order in the children's data.

Thus there were clear similarities between frequency of use by adults of particular PCF structures and frequencies with which children produced those structures in an elicitation task. This suggests that frequency in the ambient language may influence children's acquisition of particular PCF forms, to the extent that they even produce these forms in an elicitation task where the target form is modelled for them in a practice item.

Adults used a wide variety of verbs, but some verbs occurred very frequently. Certain verbs also occurred frequently in particular positions and in particular structures. Children's verb usage in the context of this elicitation task was of course much more constrained than that of adults in spontaneous speech. Many of the children's verbs specifically described aspects of the stimuli they were given. However, parallels were found between the adult and child datasets. Children made extensive use of the verb 'get' plus a range of particles; this was a common verb in adult PCFs. Perhaps more significantly, children also made quite frequent use of 'be+comparative ADJ/ NP' in the *if*-clause, describing a state in a similar way to adults' spontaneous usage. This is despite the fact that the stimulus pictures depicted actions and the results of actions. It is possible that the children may already have been aware of the adult tendency to use 'be' in PCFs to set the scene and describe states.

Adults were found to use pronominal subjects frequently in both *p* and *q* in PCFs, in line with expectations from studies of conversational corpora.

https://doi.org/10.1017/S0305000912000049 Published online by Cambridge University Press

Adults' PCFs also commonly referred to themselves as subjects (using *I*). Children's responses did not favour the first person pronoun in the same way, due to the constraints of the picture description task. However, their responses did use pronominals as subjects far more frequently than nominals. This is in line with existing research on children's narratives produced from sequences of pictures (Karmiloff-Smith, 1980; 1984; Bamberg, 1987), which shows a developmental progression in children's reference. The youngest children in these narrative studies tended to use pronouns deictically, as they would in a picture description task. However, children from around the age of six started to use pronouns to denote the main protagonist (thematic subject) in a narrative (Karmiloff-Smith, 1980; 1984; Bamberg, 1987), and referred to other participants with full NPs. Later in their development, children's reference strategies became more adult-like – pronominals were used to maintain reference, while nominals were used to switch reference (Bamberg, 1987).

While the two stimuli for this task were single pictures rather than picture sequences, children in Study 2 were being asked to do more than simply describe the picture. Story grammar (Mandler & Johnson, 1977; Stein & Glenn, 1979) gives the basic elements of narratives as setting (introducing main characters and time and location of story), one or more episodes (involving an event or problem faced by a character), an attempt (to resolve the problem), and a consequence or outcome of that attempt. Young children's narratives may or may not include all of these elements (see, e.g., Liles, 1993). PCFs describe two distinct events or states and explicitly relate them causally, and thus have something in common with rudimentary narratives; a PCF like *If she had shut the cage, the rabbit wouldn't have escaped* describes an episode and a consequence, for example. So children might be expected to employ similar strategies in their use of referring expressions in responses to pictures stimuli designed to elicit PCFs. Crosstabs comparisons were conducted between age in years and presence or absence of a pronoun subject for all four clauses. Both *p* and *q* in the 'rabbit' stimulus showed a pattern whereby the older age groups were more likely to use subjects that were not pronouns; this was a tendency for subjects in *p* ($\chi^2(5) = 15{\cdot}576$, $p < 0{\cdot}01$) and statistically significant for those in *q* ($\chi^2(5) = 17{\cdot}063$, $p < 0{\cdot}005$). As the 'rabbit' stimulus typically elicited responses that had different subjects in *p* and *q* (the girl and the rabbit), older children's preference for nominal subjects could reflect their increasing ability to indicate switches in reference (Bamberg, 1987). Given that the picture stimuli in Study 2 were always presented in the same order (rabbit, then bus), this could also explain children's increased use of full NP subjects in their 'bus' responses; there is a switch of reference between the two stimulus pictures.

https://doi.org/10.1017/S0305000912000049 Published online by Cambridge University Press

GENERAL DISCUSSION AND IMPLICATIONS FOR CHILDREN'S ACQUISITION OF THE PCF

Producing past counterfactuals is difficult for children. The explanation for this is likely to be at least partly to do with cognitive developments and constraints, such as processing load and the ability to retain and cross-reference real and imagined sets of events in memory. However, given well-designed tasks, children can understand counterfactuals long before they can produce them reliably. This suggests that linguistic complexity and cognitive complexity may be dissociable to some extent. The studies reported in this article have attempted to elucidate the contribution of linguistic complexity to the late acquisition of PCFs. In contrast to generative accounts, usage-based accounts of language acquisition are based on the assumption that children's acquisition of constructions – reliable pairings of particular syntactic forms with particular meanings – relates clearly to the occurrence of those constructions in the language surrounding children. Specifically, theoretical emphasis is placed on the frequency of occurrence of a construction in the language to which children are exposed. High token frequency (number of instances of a particular construction in the input) and low type frequency (little variation in form of the construction) mean that a construction is more likely to become 'entrenched'. By contrast, low token frequency and high type frequency mean that a construction is less likely to become entrenched, and thus will be acquired later. Despite the close relationship of form and meaning in Construction Grammar approaches (with which usage-based accounts are normally closely aligned), there is room in this model for a dissociation between cognitive and linguistic abilities. Children may well be able to understand a construction, especially if its meaning is relatively compositional (i.e. it is not too idiomatic in nature). However, in order to produce the construction accurately, they may need repeated exposure to it in order to abstract the details of form. Constructions which are highly abstract in form, such as the PCF, have higher type frequencies and are thus likely to take longer to acquire for this reason, as children are not exposed to repeated examples of the same form.

There is a great deal of variation in the structure and lexis of adults' PCFs, and they are infrequent in spontaneous adult language. The high type frequency and low token frequency of adult PCFs therefore militates against the early acquisition of the construction by children. In addition to this, PCF-like structures were used in a small number of instances to encode meanings that were not past counterfactual. Moreover, some structures (such as '*if* + past simple') were used by adults to convey meanings that may or may not be past counterfactual. This suggests that children are faced with a formidable task in identifying the form and meaning of this abstract syntactic construction. However, analysis of a large

https://doi.org/10.1017/S0305000912000049 Published online by Cambridge University Press

corpus such as the BNC can show up regularities that are not apparent 'to the naked eye', and several features of PCFs found in the BNC are hypothesized to assist children in the acquisition of the PCF construction. First, a large proportion of the PCFs in the BNC data was accounted for by a relatively small range of syntactic structures. These commonly occurring structures mirrored, to a large extent, the structures produced by children in an elicitation task, suggesting that children's acquisition may be influenced by general frequency of particular PCF structures in adult language, even if the construction as a whole appears with very low frequency. The findings from Study 1 indicated that certain verbs and subjects occurred with greater than expected frequency in adult PCFs. In Tomasello's (2003) terms, recurrent verbs and subjects may form 'item-based constructions' such as *If I had …* , *If you had …* , *it would have …* , *If I had known …* , *If it had been …* , *it would have been …* , '*If* X *had known* Y', and so on. Each of these fragments has a higher token frequency than any entire PCF, and these may then provide the basis for children's abstraction of the PCF construction. Moreover, some verbs and subjects pattern together, producing longer repeated combinations such as '*If* A'*d had* B, C *would have been* D', '*If I'd* X, *I would have* Y' and '*it would have been* ADJ *if …*'. Recurrent patterns in the input such as these could potentially assist children in building a representation of the construction as a whole. Furthermore, it is possible that the differing frequencies of particular verbs occurring in the 'canonical' PCF structure and in other common structures (such as '*If* A *had've known* (B), C *would have* D') might also help to differentiate these structures from one another for the child, and highlight structural variations that are possible within the adult model.

Study 2 showed that children's PCFs were also very varied in structure and lexical content, even in response to an elicitation task that modelled an example of the 'canonical' PCF structure and constrained children's responses to those that related to the picture stimuli. Although comparing spontaneous productions with responses to an elicitation task is not straightforward, similarities could be seen between adults' and children's patterns of usage. Children's predominant use of a small set of structures suggests the influence of adults' preferred patterns. In particular, the frequency of use of the non-standard variant *If she had of/have …* and the use of '*if*+past simple' to convey a PCF meaning are findings that it is difficult to explain outside a usage-based framework.

So far, this discussion has focused primarily on regularities in the form of PCFs. However, following Goldberg (2006), it is also possible that the meanings of lexical items which occur frequently in the PCF construction could assist children in learning PCF meanings. Pronominal subjects were common in adults' PCFs, in line with general research on conversational subjects. However, more than a quarter of the time, adults used PCFs to

https://doi.org/10.1017/S0305000912000049 Published online by Cambridge University Press

talk about themselves. Moreover, the high proportion of adult clauses containing the verb 'be' suggests that one of the primary functions of PCFs is to talk about hypothetical states, rather than (perhaps) hypothetical actions or occurrences. Adults seemed to be likely to frame *p* – the *if*-clause – as a state, rather than a description of an event. They were also likely to talk about alternative outcomes – the *would*-clause, *q* – in terms of states: what would have been, had something else taken place, rather than what would have happened. This may indicate 'core' or 'prototypical' meanings for the PCF, distinguishing it from other types of conditionals such as those with generic or predictive meanings.

Children's responses to the elicitation task also relied on a small set of verbs. Some of these were clearly related to the stimuli, but others (e.g. the use of ''be' + comparative ADJ/NP') showed a potential relationship with adult spontaneous usage. Children's elicited PCFs showed a tendency towards increasing use of full-NP subjects with increasing age, in contrast with adult spontaneous usage. However, this can be explained with reference to task effects; children's choice of subjects was in line with Bamberg's (1987) findings on the use of pronouns and full NPs to denote different types of reference.

CONCLUSIONS AND FUTURE DIRECTIONS

This article has presented evidence of substantial variation in adult PCF structure, and of regularities in adults' PCFs that are not predictable on the basis of existing grammatical descriptions of the construction. Relationships between these regularities and aspects of children's productions have been explored through reanalysis of data from an elicitation task. It was concluded that despite differences between spontaneous and elicited language data, children's productions showed the influence of features of adult usage, and that these influences are well explained with reference to a usage-based approach to language acquisition.

Future research would do well to look at children's spontaneous PCF productions using corpus data. A further elicitation study might also employ stimuli that are more in line with the 'prototypical' PCF meaning suggested in this article, i.e. representing states and settings rather than, or in combination with, actions and events.

REFERENCES

Adams, C., Cooke, R., Crutchley, A., Hesketh, A. & Reeves, D. (2001). *Assessment of comprehension and expression 6–11*. Windsor: NFER-Nelson.

Athanasiadou, A. & Dirven, R. (1997). Conditionality, hypotheticality, counterfactuality. In A. Athanasiadou and R. Dirven (eds), *On conditionals again*, 61–96. Amsterdam: John Benjamins.

https://doi.org/10.1017/S0305000912000049 Published online by Cambridge University Press

Bamberg, M. G. W. (1987). *The acquisition of narratives: Learning to use language*. Berlin: Mouton de Gruyter.
Bates, E., Bretherton, I. & Snyder, L. (1988). *From first words to grammar: Individual differences and dissociable mechanisms*. New York: Cambridge University Press.
Beck, S. R., Riggs, K. J. & Gorniak, S. L. (2010). The effect of causal chain length on counterfactual conditional reasoning. *British Journal of Developmental Psychology* **28**(3), 505–521 [doi: 10.1348/026151009X450836].
Beck, S. R., Robinson, E. J., Carroll, D. J. & Apperley, I. A. (2006). Children's thinking about counterfactuals and future hypotheticals as possibilities. *Child Development* **77**(2), 413–26 [doi:10.1111/j.1467-8624.2006.00879.x].
Bowerman, M. (1986). First steps in acquiring conditionals. In E. Traugott, C. A. Ferguson, J. Snitzer Reilly & A. ter Meulen (eds), *On conditionals*, 285–307. Cambridge: Cambridge University Press.
British National Corpus (BNC), version 3 (BNC XML Edition) (2007). Distributed by Oxford University Computing Services on behalf of the BNC Consortium [www.natcorp.ox.ac.uk/].
Burnard, L. (ed.) (2000). *The British National Corpus users reference guide*. [Retrieved from www.natcorp.ox.ac.uk/archive/worldURG/urg.pdf].
Bybee, J. & Thompson, S. A. (2000). Three frequency effects in syntax. In J. Moxley & M. Juge (eds), *Proceedings of the 23rd Annual Meeting of the Berkeley Linguistic Society*, 378–88. Berkeley, CA: Berkeley Linguistic Society.
Byrne, R. M. (1997). Cognitive processes in counterfactual thinking about what might have been. *Psychology of Learning and Motivation* **31**, 105–154 [doi:10.1016/S0079-7421(08)60501-0].
Cameron-Faulkner, T., Lieven, E. & Tomasello, M. (2003). A construction based analysis of child directed speech. *Cognitive Science* **27**, 843–73 [doi:10.1207/s15516709cog2706_2].
Carrow-Woolfolk, A. (1985). *The Test for the Auditory Comprehension of Language – Revised*. Chicago, IL: Riverside Publishing.
Clark, E. V. & Kelly, B. F. (2006). Constructions and acquisition. In E. V. Clark & B. F. Kelly (eds), *Constructions in acquisition*, 1–14. Stanford, CA: Center for the Study of Language and Information.
Crutchley, A. (2004). 'If she had of shutted the cage, the rabbit wouldn't escape': Past counterfactuals elicited from 6- to 11-year-old children. *First Language* **24**(2), 209–240 [doi:10.1177/0142723704044935].
Dancygier, B. (1998). *Conditionals and prediction*. Cambridge: Cambridge University Press.
Dancygier, B. & Sweetser, E. (2005). *Mental spaces in grammar: Conditional constructions*. Cambridge: Cambridge University Press.
Declerck, R. & Reed, S. (2001). *Conditionals: A comprehensive empirical analysis*. Berlin and New York: Mouton de Gruyter.
Dunn, L. M., Dunn, L. M., Whetton, C. & Burley, J. (1997). *British Picture Vocabulary Scale*, 2nd edn. London: NFER Nelson.
Fillmore, C. J. (1988). The Mechanisms of 'CG'. In S. Axmaker, A. Jaisser & H. Singmaster (eds), *Proceedings of the Fourteenth Annual Meeting of the Berkeley Linguistic Society*, 35–55. Berkeley: Berkeley Linguistic Society.
Fillmore, C. J. (1997a) Lecture 7: Conditional sentences. Lecture notes from Berkeley Construction Grammar website [www.icsi.berkeley.edu/~kay/bcg/lec07.html; last updated 27 February 1997; accessed 5 February 2010].
Fillmore, C. J. (1997b) Lecture 2: Idiomaticity. Lecture notes from Berkeley Construction Grammar website [www.icsi.berkeley.edu/~kay/bcg/lec02.html].
Francis, H. S., Gregory, M. L. & Michaelis, L. A. (1999). Are lexical subjects deviant? *Proceedings of the Chicago Linguistic Society* **35**, 85–97.
German, T. P. & Nichols, S. (2003). Children's counterfactual inferences about long and short causal chains. *Developmental Science* **6**, 514–23 [doi:10.1111/1467-7687.00309].
Goldberg, A. E. (2006). *Constructions at work: The nature of generalisation in language*. Oxford: Oxford University Press.

https://doi.org/10.1017/S0305000912000049 Published online by Cambridge University Press

Goldberg, A. E., Casenhiser, D. & Sethuraman, N. (2005). The role of prediction in construction-learning. *Journal of Child Language* **32**, 407–426 [doi:10.1017/S0305000904006798].

Guajardo, N. R. & Turley-Ames, K. J. (2004). Preschoolers' generation of different types of counterfactual statements and theory of mind understanding. *Cognitive Development* **19**, 53–80 [doi:10.1016/j.cogdev.2003.09.002].

Huttenlocher, J., Smiley, P. & Charney, R. (1983). Emergence of action categories in the child: Evidence from verb meanings. *Psychological Review* **90**, 72–93 [doi:10.1037//0033-295X.90.1.72].

Karmiloff-Smith, A. (1980). Psychological processes underlying pronominalization and non-pronominalization in children's connected discourse. In J. Kreiman & E. Ojedo (eds), *Papers from the Parasession on Pronouns and Anaphora*, 231–50. Chicago: Chicago Linguistic Society.

Karmiloff-Smith, A. (1984). Children's problem solving. In M. E. Lamb, A. L. Brown & B. Rogoff (eds), *Advances in developmental psychology*, vol. III, 39–90. Mahwah, NJ: Erlbaum.

Kay, P. & Fillmore, C. J. (1999). Grammatical constructions and linguistic generalizations: The What's X Doing Y? construction. *Language* **75**(1), 1–33 [doi:10.2307/417472].

Krippendorff, K. (2004). *Content analysis: An introduction to its methodology*, 2nd edn. Thousand Oaks, CA: Sage.

Krippendorff, K. (2007). Computing Krippendorff's alpha-reliability. [Retrieved from www.asc.upenn.edu/usr/krippendorff/dogs.html].

Kuczaj, S. A. & Daly, M. J. (1979). The development of hypothetical reference in the speech of young children. *Journal of Child Language* **6**, 563–79.

Liles, B. Z. (1993). Narrative discourse in children with language disorders and children with normal language: A critical review of the literature. *Journal of Speech and Hearing Research* **36**, 868–82.

Lombard, M., Snyder-Duch, J. & Bracken, C. C. (2002). Content analysis in mass communication: Assessment and reporting of intercoder reliability. *Human Communication Research* **28**, 587–604 [doi:10.1093/hcr/28.4.587].

Mandler, J. M., & Johnson, N. S. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive Psychology* **9**, 111–51 [doi:10.1016/0010-0285(77)90006-8].

Perner, J., Sprung, M. & Steinkogler, B. (2004). Counterfactual conditionals and false belief: A developmental dissociation. *Cognitive Development* **19**, 179–201 [doi:10.1016/j.cogdev.2003.12.001].

Reilly, J. S. (1982). The acquisition of conditionals in English. Unpublished PhD dissertation, University of California, Los Angeles.

Reilly, J. S. (1983). What are conditionals for? *Papers and Reports on Child Language Development* **22**, 124–32.

Reilly, J. S. (1986). The acquisition of temporals and conditionals. In E. Traugott, C. A. Ferguson, J. Snitzer Reilly & A. ter Meulen (eds), *On conditionals*, 309–331. Cambridge: Cambridge University Press.

Riggs, K. J., Peterson, D. M., Robinson, E. J. & Mitchell, P. (1998). Are errors in false belief tasks symptomatic of a broader difficulty with counterfactuality? *Cognitive Development* **13**, 73–90 [doi:10.1016/S0885-2014(98)90021-1].

Robinson, E. J. & Beck, S. (2001). What is difficult about counterfactual reasoning? In K. J. Riggs & P. Mitchell (eds), *Children's reasoning and the mind*, 101–120. Hove: Psychology Press.

Sag, I. A. (to appear). Sign-based construction grammar: An informal synopsis. In H. Boas and I. A. Sag (eds), *Sign-based construction grammar*. Stanford: CSLI Publications [Retrieved from http://lingo.stanford.edu/sag/papers/theo-syno.pdf].

Schachter, J. C. (1971). Presupposition and counterfactual conditional sentences. Unpublished PhD dissertation, University of California, Los Angeles.

Schönefeld, D. (2006). Constructions. *Constructions* Special Volume **1**(1) [www.constructions-online.de/articles/specvol1/667].

https://doi.org/10.1017/S0305000912000049 Published online by Cambridge University Press

Scott, M. (2004). WordSmith Tools v. 4.0. Oxford: Oxford University Press [Retrieved from www.lexically.net/wordsmith].
Stein, N. L. & Glenn, C. G. (1979). An analysis of story comprehension in elementary school children. In R. O. Freedle (ed.), *New directions in discourse processing*, vol. 2, 53–120. Norwood, NJ: Ablex.
Thomson, A. J. & Martinet, A. V. (1986). *A practical English grammar*, 4th edn. Oxford: Oxford University Press.
Tomasello, M. (1992). *First verbs*. Cambridge: Cambridge University Press.
Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

APPENDIX

List of all PCF structures in the BNC with examples, frequencies and percentages

Structure	Example	Freq	%
if + had + lexical verb, would + have/of + lexical verb	but if he had sent him to the eye hospital he'd have waited two years.	135	34·7
if + had + have/of + lexical verb, would + have/of + lexical verb	mum said if they'd have come erm, she'd have put them in the bathroom.	91	23·4
if + past simple, would + have/of + lexical verb	if they took my wages into consideration they would have let us buy next door even.	53	13·6
if + had + lexical verb, would + lexical verb	He'd be on his back if he 'd had that amount of drink.	23	5·9
if + had + lexical verb, present simple	What happens if Labour had got one more vote than Conservative?	10	2·6
if + present simple, would + have/of + lexical verb	if there is an easy way of doing it the Conservatives would have found out and done it.	10	2·6
if + had + have/of + lexical verb, would + lexical verb	if you'd of stayed with him you'd have your hands full.	6	1·5
if + would have/of + lexical verb, would + have/of + lexical verb	if we would have gone somewhere else and had coffee and a cake it would of cost you, one fifty each.	6	1·5
if + past continuous, would + have/of + lexical verb	if he still wasn't speaking to me properly I wouldn't have bought that off him!	5	1·3
if + had + lexical verb, had + lexical verb	if I'd got here a minute quicker I'd got you some chocolate biscuits you could of been having with that cup of tea	5	1·3
if + modal + lexical verb, would + have/of + lexical verb	if the council could, could've come to some agreement and put in there, it would of made it right for us.	4	1·0
if + had + have/of + lexical verb, present simple	Can't be him, if you'd have paid five quid a ticket.	4	1·0
if + had + have/of + lexical verb, past simple	that was a waste of their time if they'd have gone round there to view them, weren't it?	4	1·0
if + had + lexical verb, would + have/of + lexical verb	I would have understand if he'd dived and got out their way and the wind just blew it the other end.	2	0·5
if + had + lexical verb, have + lexical verb	I can't ever remember reading James. If I had, I've forgotten about it.	2	0·5

https://doi.org/10.1017/S0305000912000049 Published online by Cambridge University Press

(*Cont.*)

Structure	Example	Freq	%
if + had + lexical verb, would + present continuous	if I hadn't got them, I'd be thinking now was it on the card I haven't got you know.	2	0·5
if + had + have/of + lexical verb, past continuous	we thought it was a boy that had written and if it had've been we were gonna get the police because she's under age you see.	2	0·5
if + had + have/of + lexical verb, would + lexical verb	if you'd have got eight, nine, you'd been in!	2	0·5
if + had + past continuous, would + have/of + lexical verb	if our Margaret had er been working, I honestly think she would have left Pete.	2	0·5
if + have + lexical verb, would + have/of + lexical verb	And there was not a lot unless the people had had three bits of beef. If they've had three bits of beef then there would have been.	2	0·5
if + lexical verb, would + have/of + lexical verb	he wouldn't of told it if it gone on this.	1	0·3
if + had + have/of + past continuous, would + have/of + lexical verb	if we hadn't actually have been starting the work, heaven knows how long it would have been before we'd actually got the invoices paid!	1	0·3
if + had have/of + catenative, would + lexical verb	If he'd have kept hitting it like I h I hit mine no way.	1	0·3
if + had + lexical verb, will + lexical verb	if she hadn't had anything she'll have something when she comes back.	1	0·3
if + had + lexical verb, past simple	If you're if you hadn't earned the money well you didn't go into it in detail and get the time and motion study man.	1	0·3
if + had + lexical verb, present continuous	even if I had lost weight this week it's only going to be a couple of pounds.	1	0·3
if + had + lexical verb, to + lexical verb	if you hadn't got it by the Tuesday, to let us know on the Wednesday aye, next week mm.	1	0·3
if + had + lexical verb, would + have/of + past continuous	Whereas if it had been standing room, we really would have been standing.	1	0·3
if + had + lexical verb, would + lexical verb	if Labour'd got in they'd of er, your dad said he'd er probably got a rise.	1	0·3
if + had + lexical verb, lexical verb	A: if you'd had that B: Mm A: at mark, that twenty six quid B: Yeah A: that be B: More fifty quid.	1	0·3
if + had + lexical verb, would + past continuous	If things had turned out differently I'd been driving you down here.	1	0·3
if + had + have/of + lexical verb, present continuous	I'm having this afternoon if he'd have come here.	1	0·3
if + had + have/of + lexical verb, would + present continuous	Well if we hadn't of had that shower I'd be working on me hedge now	1	0·3
if + had + have/of + lexical verb, will + be + continuous	he'll be going in, in with bloody mob if I'd have done it.	1	0·3
if + had + past continuous, past simple	if he'd been paying quite a bit next time he came in the shop they was all round him trying to persuade him to buy a mac.	1	0·3

https://doi.org/10.1017/S0305000912000049 Published online by Cambridge University Press

(*Cont.*)

Structure	Example	Freq	%
if + present simple, would + have/of + lexical verb	because I wouldn't of understand any of that if I had that and this really explains a lot of it.	1	0·3
if + past simple, would + have/of + lexical verb	if I had any children I'd have give to them.	1	0·3
if + would have/of + lexical verb, present simple	Well what about if England had won? If Holland would've still beaten Poland? That's the end of the game, that's the end of it.	1	0·3
?	I would have asked Chris if he is pay for his erm.	1	0·3
Total		389	100·0

https://doi.org/10.1017/S0305000912000049 Published online by Cambridge University Press